DOCUMENT RESUME

ED 383 806 UD 030 477

TITLE	Indicators of Children's Well-Being: Conference Papers. Child Health, Education, and Economic Security. Volume II. Special Report Series. Special Report Number 60b.
INSTITUTION	Wisconsin Univ., Madison. Inst. for Research on Poverty.
PUB DATE	May 95
NOTE	317p.; For Volumes I and III, see UD 030 476 and UD 030 478. Papers presented at a conference on "Indicators of Children's Well-Being" (Bethesda, MD, November 17-18, 1994).
AVAILABLE FROM	Institute for Research on Poverty, 1180 Observatory Drive, Madison, WI 53706 ($10).
PUB TYPE	Collected Works - Conference Proceedings (021) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	*Academic Achievement; Adolescents; *Child Health; Children; Child Welfare; Dropouts; *Economic Factors; Elementary Secondary Education; Evaluation Methods; Low Income Groups; Poverty; *Preschool Education; *School Readiness; Vocational Education
IDENTIFIERS	*Indicators

ABSTRACT

This volume groups papers discussing indicators of children's well-being into three areas: child health, education, and economic security. Papers concerning child health include: (1) "Prenatal and Infant Health Indicators" (Paula Lantz and Melissa Partin); (2) "Health Indicators for Preschool Children (Ages 1-4)" (Barbara L. Wolfe and James Sears); (3) "Health Indicators for Preadolescent School-age Children" (Barbara Starfield); (4) "Adolescent Health Indicators" (Arthur B. Elster); and (5) "Indicators for Infant, Child, Preadolescent and Adolescent Health: Discussion (Michael D. Resnick). The second section, Education, includes the following papers: (1) "Indicators for School Readiness, Schooling, and Child Care in Early to Middle Childhood" (Deborah Phillips and John Love); (2) "Indicators of High School Dropout" (Robert M. Hauser); (3) "Postsecondary and Vocational Education: Keeping Track of the College Track" (Thomas J. Kane); (4) "Indicators of Educational Achievement" (Daniel Koretz); and (5) a Summary of the four papers on education indicators (Richard J. Murnane). The last section, Economic Security, includes the following papers: (1) "Income, Employment, and the Support of Children" (Susan E. Mayer); (2) "Longitudinal Indicators of Children's Poverty and Dependence" (Greg J. Duncan); and (3) "Parental Employment and Children" (Judith R. Smith, Jeanne Brooks-Gunn, and Aurora P. Jackson). Most papers contain references. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Bert Penn
Inst. Research Poverty
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

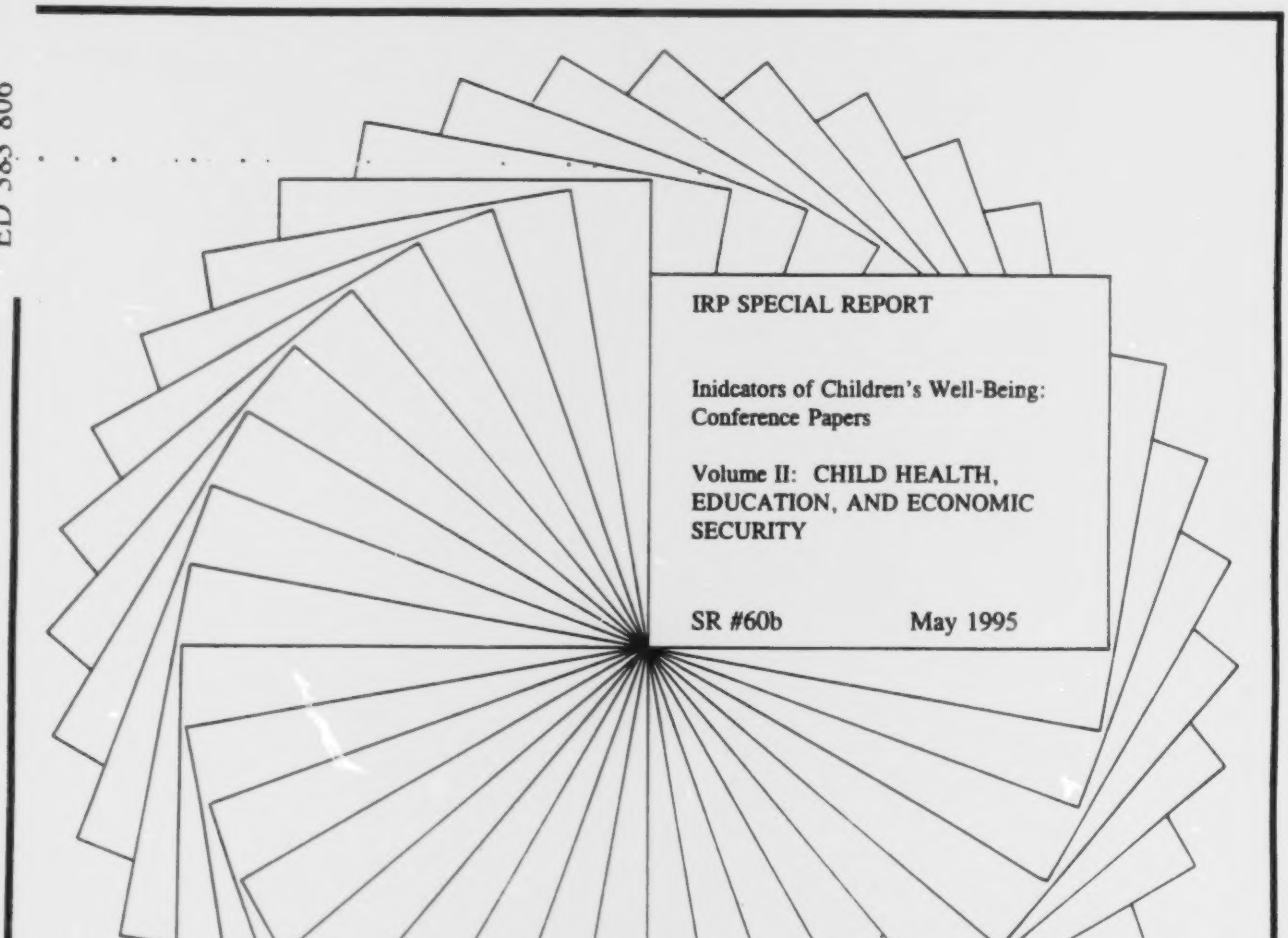
University of Wisconsin-Madison

Institute for Research on Poverty

Special Report Series

ED 383 806

UD 030 477



IRP SPECIAL REPORT

Inidcators of Children's Well-Being:
Conference Papers

Volume II: CHILD HEALTH,
EDUCATION, AND ECONOMIC
SECURITY

SR #60b May 1995

BEST COPY AVAILABLE

Institute for Research on Poverty
Special Report no. 60b

Indicators of Children's Well-Being:
Conference Papers

Volume II:
Child Health, Education, and Economic Security

May 1995

Contents

IRP Special Report 60b
Indicators of Children's Well-Being: Conference Papers
Volume II: Child Health, Education, and Economic Security

Child Health

1. Prenatal and Infant Health Indicators 3
Paula Lantz and Melissa Partin

2. Health Indicators for Preschool School-Age Children 33
Barbara L. Wolfe and James Sears

3. Health Indicators for Preadolescent Children 49
Barbara Starfield

4. Adolescent Health Indicators 71
Arthur B. Elster

Discussant's Comments 83
Michael Resnick

Education

1. Indicators for School Readiness, Schooling, and Child Care in Early to Middle Childhood 91
Deborah Phillips and John Love

2. Indicators of High School Dropout 117
Robert H. Hauser

3. Postsecondary and Vocational Education: Keeping Track of the College Track 153
Thomas J. Kane

4. Indicators of Educational Achievement 181
Daniel Koretz

Discussant's Comments 203
Richard Murnane

Economic Security

1. Income, Employment, and the Support of Children 225
Susan E. Mayer

2. Longitudinal Indicators of Children's Poverty and Dependence 263
Greg J. Duncan

3. Parental Employment and Children 279
Judith R. Smith, Jeanne Brooks-Gunn, and Aurora P. Jackson

CHILD HEALTH

PRENATAL AND INFANT HEALTH INDICATORS

Paula Lantz, Ph.D.
University of Michigan School of Public Health
Ann Arbor, Michigan

Melissa Partin, Ph.D.
Minnesota Department of Health
Minneapolis, Minnesota

October, 1994

Prepared for the Conference on Indicators of Children's Well-Being, Rockville, Maryland, November 17-18, 1994.

The alphabetical listing of the authors reflects their equal contribution to the paper.

I. INTRODUCTION

During 1992, over 4 million babies were born in the United States (NCHS, 1993). The health and well-being of infants and their mothers is of critical importance to our society. The health status of the mother during the prenatal period is inextricably related to pregnancy outcome and to the baby's health during infancy (i.e. the first year of life). In turn, health status during infancy has a crucial impact on the future health and development of growing children.

It is commonly accepted that the foundation for all aspects of life (physical, social and emotional) is laid during its earliest stages. Children are indeed the future; and their well-being before birth and during infancy are of great importance to that future. In addition, the health status and well-being of pregnant women and their infants says much about a society, for healthy children imply a healthy society. Many statistics and other indicators of prenatal and infant health are part of common, everyday language and are used to compare health conditions for mothers and infants across population subgroups and across nations as well. For example, the fact that the rate of infant mortality in the United States ranks well behind most other developed countries and some underdeveloped countries is well known and a source of great national concern (Haub and Yanagishita, 1991).

The purpose of this paper is to discuss aggregate information and indicators that can be used to assess the health and well-being of children during the prenatal period and infancy. We begin by presenting a comprehensive list of prenatal and infant health indicators, and discussing the major sources of information on these indicators. We next identify a set of three priority or key indicators from the comprehensive list of indicators and provide a justification for their selection. We then evaluate the three key indicators with regard to their availability, quality, and usefulness for measuring prenatal and infant health status. As part of this discussion, we present an assessment of the strengths and limitations of each key indicator, and provide recommendations for improved data collection during the coming decade. We also present a brief discussion of additional prenatal and infant health indicators that were not selected as priority indicators but nonetheless are important and deserve mention.

II. PRENATAL AND INFANT HEALTH INDICATORS AND PRIMARY DATA SOURCES

To compile a list of important indicators of prenatal and infant well-being, several sources of information were used. These sources included: 1) scientific literature from medicine, public health and the social sciences; 2) the United States Public Health Service's goals and objectives for the Year 2000, as outlined in *Healthy People 2000: National Health Promotion and Disease Prevention Objectives* (hereafter referred to as Year 2000 objectives) (1990); and 3) materials and reports prepared by child advocacy groups such as the Children's Defense Fund (1992, 1994), and the Carnegie Task Force on Meeting the Needs of Our Youngest Children (1994).

Table 1 displays a list of important direct indicators of prenatal and infant well-being (i.e., those measures which describe various aspects of the health status and well-being of mothers and fetuses during the prenatal period and of babies during infancy). The list of indicators presented in the table is by no means exhaustive. Additional indicators and measures have been used to assess aspects of prenatal and infant well-being and to identify health risks during these time periods. The indicators included in our list were selected based on the following criteria: 1) each indicator definition is clear, objective and measurable; 2) each indicator's definition remains consistent across population subgroups and has remained stable over time; and 3) each indicator has meaning for and is generally understood by the public. As a group, the indicators assess well-being across a wide range of outcomes, processes and behaviors, and include both positive and negative measures of well-being. The first three indicators listed in Table 1 (measures of infant mortality, low birth weight, and prenatal care utilization) were selected as the most important or key indicators of prenatal and infant health. These priority indicators are described in detail below in Section III. Other indicators of prenatal and infant health are briefly discussed in Section IV.

The majority of the information and measures used to assess prenatal and infant health in the United States are derived from three types of data: 1) vital registration data; 2) medical records data (including patient medical charts, patient laboratory and procedure records, patient billing records, and hospital discharge data); and 3) survey research data. All of these data sources are available in a variety of formats (including in combined forms) at the national, state and local level (NCHS, 1993; Gable, 1990). Table 2 provides a list of specific sources of data for each of these categories.

Although our focus primarily is on national data sources, it is important to emphasize that many states also have excellent data sources for assessing prenatal and infant health at the state, county or other local level. For example, many state centers for health statistics link their birth and death certificates to produce state-specific information on birth outcomes. In addition, most states aggregate hospital discharge information that is used to compare data on perinatal hospital diagnoses, lengths of stay, treatment costs, and outcomes across geographic regions and subpopulations within the states.

Each of the main types of data used for prenatal and infant health assessments has strengths and weaknesses. Several measures of prenatal and infant health (including our three key indicators) can be attained from vital registration data alone. The strengths of vital record data are that coverage for all births and deaths is nearly complete, data collection methods and forms are similar across geographic regions and sociodemographic groups, and much work has already been invested in assessing and improving data quality. A main concern regarding vital registration data is that the quality of some of the data elements on birth and death certificates is suspect. Studies have found quality problems associated with a variety of data elements, including length of gestation, obstetric complications, medical interventions during pregnancy, and reports of the use of alcohol and other drugs during pregnancy (Carver et al., 1993; David, 1980; NCHS, 1985; Frost et al., 1984; Kramer et al., 1988; Oates and Forrest; 1984; Parrish et al., 1993; Piper et al., 1993). An additional concern is that the turn-around time for indicator availability is lengthy. For example, as of late 1994, the most recent national statistics on infant mortality were for 1991.

Although medical records data can provide useful information (especially at the local level) that cannot be obtained through vital records or self-report survey data, this type of data generally is difficult to obtain. The approval of hospital or clinic officials and/or institutional review boards is often required before patient record information is released. Furthermore, since patient charts and other record keeping systems greatly differ across institutions, it is difficult to combine data collected from a number of medical settings. The costs involved with aggregating medical records data tend to be high, especially if the effort involves chart abstraction. In addition, in most settings it is difficult if not impossible to link maternal prenatal care records, hospital records for the mother and newborn, and subsequent pediatric records for the child, making research on the association of maternal characteristics and prenatal care with birth outcomes and infant health challenging.

Surveys often provide interesting and useful information that is not captured in vital registration data or in medical records. The main problem with survey results, however, is that they most often are based on self-reported data from the selected respondents or their informants. The main sources of error in self reports stem from **recall loss** (i.e. the respondent does not recall events and situations or their timing and thus does not report accurately) and **intentional bias** (i.e. the respondent gives false or incomplete information on purpose) (Harlow and Linet; 1989; Hewson and Bennet, 1987; Tilley et al., 1985).

Combining information from different data sources can often yield better results (Hexter et al, 1990). For example, at the state level, information from a hospital discharge survey can be combined with vital records information to produce a richer data source on infant and maternal health at the of delivery (Parrish et al., 1993). At the national level, an example of a combined data source is the 1980 National Natality Survey, which includes information from birth certificates, medical records, and maternal self report from a survey questionnaire.

III. INDICATORS OF HIGHEST PRIORITY FOR PRENATAL AND INFANT HEALTH

Three of the indicators from the list of direct indicators (Table 1) were selected as being the most important or of the highest priority for assessing prenatal and infant well-being in the United States. These key indicators include: 1) measures of infant mortality; 2) measures of low birth weight; and 3) measures of prenatal care utilization. These key indicators were selected as being of highest priority for several reasons. The indicators refer to areas of critical importance to health and well-being during the prenatal and infant periods. In addition, the indicators are meaningful across subpopulation groups and across cultures and nations. Comparable data for all three indicators are available at the national, state, and local levels, and can be broken down by race, maternal age and other factors. Finally, data collection methods are similar in all states, overseen by the National Vital Statistics System (NCHS, 1993), and data collection procedures have remained relatively stable over time.

The three indicators chosen reflect societal norms and goals. One cannot imagine dissent from the opinions that pregnant women deserve adequate prenatal care, that babies ought to be born mature and healthy, and that babies ought to survive through infancy and beyond. In addition, these indicators are already widely reported and used to assess prenatal and infant health in a variety of formats. These formats include national surveillance data prepared by government agencies, the Year 2000 health promotion/disease prevention objectives, the reports and materials prepared by child advocacy groups, and academic maternal and child health research. In addition, these indicators are generally understood and appreciated by the media and the lay public. Thus, there appears to be widespread agreement that measures of infant mortality, low birth weight and prenatal care are important and meaningful indicators of prenatal and infant health.

The selected key indicators are certainly related to each other. Use of prenatal care is a strong predictor of both birth weight and infant survival (Eisner et al., 1979; Greenberg, 1983; Leveno et al;, 1985). Low birth weight, in turn, is a major determinant of infant mortality and the major cause of neonatal death (McCormick, 1985). Also important, however, is the fact that these indicators are strongly associated with many other aspects and indicators of infant health. An infant's exposure to prenatal care and his/her birth weight are not only related to survival through infancy but also to numerous other aspects of well-being during the first year of life, such as physical development and morbidity.

In the sections below, we describe in detail the state of each of the three priority indicators and the various ways in which the indicators are produced and used. We also address three questions for each indicator: 1) how is the indicator currently measured?; 2) how should the indicator be produced?; and 3) how can improved measures be produced over the next decade?

A. Measures of Infant Mortality

How is Infant Mortality Currently Measured?

Definitions: Infant mortality is a measure of infants' survivability through the first year of life. The infant mortality rate (IMR) is a ratio of the number of deaths to children under the age of one compared to the number of live births during a specified time period (usually a year). The crude or conventional IMR can be defined as follows (Shryock and Siegel, 1976):

$$\frac{\text{Deaths to children} < 1 \text{ year of age during the year}}{\text{Births during the year}} \times 1{,}000$$

The majority of infants who die during the first year do so during the first weeks of life (McCormick, 1985). In addition, the causes of death for those babies dying very early in infancy differ significantly from those dying later during this time period. Thus, the overall infant mortality rate can be broken down into two component parts: the neonatal mortality rate and the post-neonatal mortality rate. The **neonatal mortality rate** measures the level of death during the first four weeks of infancy (i.e. less than 28 days of age). The **post-**

neonatal mortality rate measures the level of death after the first month (i.e. between 28 and 364 days of age). The neonatal mortality rate is used as a measure of endogenous mortality, since the majority of neonatal deaths are due to causes that are congenital or endogenous to the mother and/or baby (i.e. prematurity or congenital defects). Alternatively, the post-neonatal mortality rate has been used as a measure of exogenous mortality, since a higher proportion of post-neonatal deaths are due to causes of death which are external to the mother and child (i.e. nonintentional injury, respiratory infections). However, as improvements in perinatal medicine have extended the survival time of infants born very ill, the assumption that post-neonatal deaths are primarily due to exogenous causes has become less valid.

Cause-of-death-specific neonatal and post-neonatal mortality rates by race and ethnicity provide useful information for assessing the racial/ethnic differences in the timing and causes of infant death. Also useful are infant mortality rates by birth weight categories and gestational age in weeks. Trends in both of these indicators are useful for many types of assessments, including the tracking of improvements in perinatal medicine.

Infant mortality is a relatively rare event. Although infant death may be rare in developed countries, the infant mortality rate is a widely-used indicator of development and the overall health of a society, since it reflects medical technology, hygiene and sanitation systems, and the availability and use of both preventive and clinical health services. Several developing nations have infant mortality rates of over 100, indicating that over 10 percent of babies born do not survive infancy (International Bank for Reconstruction and Development, 1984). The comparison of infant mortality rates across more developed countries, however, can also be illustrative. As mentioned previously, the United States has one of the highest infant mortality rates in the developed world (Schieber et al., 1991). According to 1987 data, the infant mortality rate in the United States was higher than 21 other developed nations for which comparable data were available (Haub and Yanagishita, 1991). The overall rate for the United States was 10.1 deaths per 1,000 live births, compared with 5.0 for Japan, 6.1 for Sweden, 7.3 for Canada, and 9.0 for Spain. In addition, sociodemographic differentials in the infant mortality rate within a country reflect the extent to which health resources, the prevalence of risk behaviors, and measures of health status are differentially distributed in a society. For example, the infant mortality rate for blacks has persisted at a rate of at least twice that of whites since 1915, when rates by race were compared for the first time. In 1991, the infant mortality rate for white infants was 7.3, while the rate for black infants was 17.6. Thus, although infant mortality is a rare event in the United States, it is an elucidating and quite useful indicator of overall infant health and of sociodemographic differences in health and resource distribution.

Data Production: The data used to compute measures of infant mortality (and other indicators of prenatal and infant health) come from vital records. The registration of births and deaths is required by law in all states and territories of the United States. All states, therefore, have vital registration data on births and deaths that can be assessed at the state, county or municipal level. National data on births and infant deaths are available through the National Vital Statistics System, a data collection effort of the National Center for Health Statistics (NCHS) (Perrin, 1974). Through this system, NCHS collects and publishes data on births, deaths, marriages, and divorces in the United States. The Division of Vital Statistics receives birth and death information from the registration offices of all 50 states, New York City, the District of Columbia, Puerto Rico, the Virgin Islands, and Guam.

Since 1933, geographic coverage for birth and death registration has been complete (*NCHS*, 1993). At the present time, all 50 states and the District of Columbia participate in the Vital Statistics Cooperative Program. Through this NCHS program, all birth and death records (rather than a sample) are sent to NCHS on computer tapes. NCHS cooperates with states to obtain the data necessary to produce national vital statistics, and to improve the timeliness and quality of this health data (NCHS, 1991). It is generally accepted that at least 99 percent of all live births and deaths are captured in the national vital records system (NCHS, 1993; NCHS, 1991; Frost et al., 1982; U.S. National Office of Vital Statistics, 1978).

United States standard certificates for live births, deaths, marriages, and divorces, and standard reports for induced termination of pregnancy and fetal deaths are periodically revised (in approximately 10-year cycles). The standard certificates/reports represent the minimum data needed to produce comparable national, state, and local vital statistics. The most recent revisions were implemented in 1989 (NCHS, 1991; Freedman et al., 1988; MacFarlane, 1989). It is believed that these changes will improve the quality of the data gathered and will provide new and increased opportunities for research on birth outcomes (Taffel et al., 1989;, Freedman et al., 1988; Luke and Keith, 1991). Nearly all registration areas for which NCHS publishes data were using the revised standard forms by January 1, 1989 (NCHS, 1991).

In addition to housing separate files containing annual birth and death certificate information, NCHS also links vital records for research on infant mortality. The national linked file of live births and infant deaths is comprised of linked birth and death records for infants born in a given calendar year who died before their first birthday (NCHS, 1993). Two years worth of vital statistics data are required for the construction of the linked file, since infant deaths can occur during the year of birth and the year after. The match completeness for the linked files is high (i.e. 98% for the 1983-1987 files) (NCHS, 1993). This national file can be used to assess prenatal and infant health at the state and local level as well.

State and county infant mortality statistics typically are produced on an annual basis and disseminated by state centers for health statistics. National infant mortality statistics are also produced on an annual basis and are published in a variety of places including various NCHS reports, the *Health, United States* series, and the *Morbidity and Mortality Weekly Report*. The turn-around time for the production of annual infant mortality statistics is generally between two and four years.

Infant mortality statistics can be produced with information from infant death certificates and a count of the number of live births during the same time period. With this information, infant morality rates by cause of death and by timing of death (neonatal versus post neonatal) can be computed. If the number of live births is available by race and ethnicity, race/ethnic specific infant mortality rates can also be produced. When death certificate information is combined with data from birth certificates, infant mortality rates can be assessed by birth weight, timing and use of prenatal care, and other relevant factors on the birth certificate. Thus, files which link birth and death certificate data provide a rich source for producing measures of infant mortality (NCHS, 1986; Zahniser et al., 1987).

Data Quality: Birth certificate coverage of live births and death certificate coverage of infant deaths are believed to be quite high (U.S. National Office of Vital Statistics, 1978; Frost, 1982; Kleinman, 1986; Lambert and Strauss, 1987). Nevertheless, concerns regarding the underreporting of fetal, perinatal and infant deaths have been documented (Kleinman, 1986; David, 1986; Williams et al., 1986). Even a small number of unreported out-of-hospital births and deaths could have a substantial impact on mortality rates for racial, ethnic or other subpopulations (Kleinman, 1986). The main quality concerns regarding infant mortality indicators, however, involve cause of death information and race identification on death certificates.

Several studies have found discrepancies in cause of death information when autopsy results are compared with the cause of death codes on corresponding death certificates (Kircher et al., 1985; Schottenfeld et al., 1982; Carter, 1985). With regard to infant deaths, it is believed that cause of death statistics from death certificates underestimate deaths due to a number of diseases and underlying conditions, including congenital anomalies (Minton and Seegmiller, 1986), child abuse and neglect (McClain et al., 1993), and the impact of a short gestation (Carver et al., 1993). The problems associated with cause of death information on death certificates are believed to be related to several factors. First, the immediate function of the death certificate is legal (i.e. to permit transfer of the body and to initiative appropriate claims). Thus, the document is usually completed as quickly as possible and is rarely edited or modified by autopsy or other subsequent findings (Kircher et al., 1985;, Carter, 1985; Buetow, 1992) Second, the majority of physicians have no training in the purpose and process of death certification (Comstock and Markush, 1986). Third, physicians are not routinely queried about incomplete diagnoses, unlikely sequences, or missing information. (Comstock and Markush, 1986; Rosenberg, 1989) Finally, with regard to the underestimation of infant deaths due to short gestation, it has been argued that biases in World Health Organization (WHO) selection rules allow other immediate causes of death to have a higher priority over short gestation (Carver et al., 1993).

There is empirical evidence that there are data quality problems associated with the coding of race and ethnicity on birth and death certificates. A study of matched birth and death certificates over a 13-year period in Oklahoma revealed that 28% of infants born American Indian were classified as another race (typically as white) on death certificates (Kennedy and Deapen, 1991). Corrections in the coding of race at death nearly doubled the infant mortality rate of American Indians. Similarly, a national study of 1983-1985 birth and death certificates found great inconsistencies in the coding of race and ethnicity (Hahn et al., 1992). Overall, 3.7% of the infant death certificates had a different race or ethnicity code than birth certificates. The majority (87%) of infants classified differently at death were coded as being white. Hahn and colleagues (1992) also found that consistent coding of race/ethnicity at birth and death produces infant mortality rates that are 2.1% lower for whites, 3.2% higher for blacks, 46.9% higher for American Indians, 33.3% higher for Chinese, 48.8% higher for Japanese, 78.7% higher for Filipinos, and 8.9% higher for Hispanics. Disparities in the coding of race on birth and death certificates cast doubt on the accuracy of race-specific infant mortality statistics, especially those for minorities. Improvements in the coding of race and ethnicity on birth and death certificates need to be made (Kennedy and Deapen, 1991; Hahn et al., 1992; Nakamura et al., 1991; Becerra et al., 1991). In addition, however, the issues of multiracism and how definitions of race and ethnicity have changed over time also need to be acknowledged and addressed if statistical indicators involving race are to be meaningful (Wright, 1994).

Since much research on infant mortality is conducted on files which link birth and death certificates, the quality of information on birth certificates is also of importance. Studies have found birth certificate data on birth weight, APGAR scores, maternal education, and other sociodemographic variables to be of relatively high quality (Brunskill, 1990; David, 1980; NCHS, 1985; Jepson et al., 1991; Oates and Forrest, 1984; Piper et al., 1993; Querec, 1980). There is some evidence, however, that birth certificate data on gestational age, prenatal care utilization, maternal health complications, and congenital anomalies and abnormal conditions of the newborn do have some problems related to quality (Alexander et al., 1991; Alexander et al., 1990; Carver et al., 1993; David, 1980; NCHS, 1985; Frost et al., 1984; Hexter et al., 1990; Kramer, 1988; Parrish et al., 1989; Piper et al., 1993; Querec, 1980). It was hoped, based on previous studies, that the 1989 revision of the Standard Certificate of Birth would improve the quality of these items through the provision of the checkbox format (Frost et al., 1984; NCHS, 1991; Taffel et al., 1989) Quality issues related to birth certificate items will be discussed in greater detail below.

Methods of Data Collection: Infant mortality rates are produced from vital registration data. This does not mean, however, that other sources of information are not useful or essential to the study of infant mortality. Information from alternate sources can augment the data available through the vital records system. For example, linking vital records with hospital discharge information can provide data on the costs associated with caring for premature infants who eventually die (Hexter et al., 1990; Parrish et al., 1993). In addition, information from survey questionnaires provides rich opportunities for researchers to investigate explanations for observed sociodemographic differentials in infant death. For example, Geronimus et al. (1991) used data from the National Health and Nutrition Examination Survey to investigate the hypothesis that racial differences in the age-specific prevalence of health conditions associated with maternal pregnancy complications (i.e. hypertension) may explain some portion of long-observed racial differences in pregnancy outcome.

How Should Infant Mortality Rates be Produced?

The infant mortality rate is an indicator of the incidence or occurrence of infant death, rather than an indicator of the prevalence or preponderance of a specific health condition. The unit of analysis is individual infants (i.e. the number of infant deaths per live births in a year), as opposed to a unit of analysis involving the mother or family. As mentioned above, national, state and local rates typically are produced for a calendar year. The production of annual infant mortality statistics seems sufficient, and we see no reason to increase or decrease this rate of production.

It is important to emphasize that infant mortality rates are not probability measures (Shryock and Siegel, 1976). These rates reflect the number of deaths during a year relative to the number of live births. To the extent that the babies dying during a year were born in the previous year, the infant mortality rate is not a probability. It is an approximation of the probability of overall mortality and neonatal mortality; it is less of an approximation of a probability for post-neonatal mortality. Linked birth and death certificate files (at both the state and the national level) provide ample opportunities for researchers and others to produce actual probabilities when the need arises.

The distinction between neonatal and post-neonatal mortality continues to be important, thus we recommend that infant mortality rates for these two different age groups continue to be produced. In addition, infant mortality data for population subgroups are very important. As mentioned earlier, sociodemographic differentials (both levels and trends) are very elucidating and are considered to reflect differences in life style, access to medical care, and health-related knowledge, attitudes, and practices. We recommend that, at a minimum, national and state infant mortality rates be produced by race/ethnicity and maternal age. In addition, cause-specific infant mortality rates should also be produced on an annual basis. In all cases, new annual statistics should be compared with previous years to identify trends in both levels and patterns of infant mortality.

Analyses of trends in infant mortality rates should include adjustments for several other concurrent trends. The major demographic trends that need to be considered include: 1) changes in the distribution of maternal age by race; 2) changes in the distribution of maternal parity by race; 3) changes in the birth rate by race. All of these demographic changes can have an impact upon crude infant mortality rates and/or their interpretation. For example, since blacks have a higher rate of infant mortality than whites, the overall infant mortality rate is influenced by the proportion of births to black mothers. In addition, analyses of trends in infant mortality rates have attempted to adjust for changes in maternal behavior and social policy in addition to changes in demographics. For example, attempts have been made to adjust or explain the widening of the black/white infant mortality gap by incorporating information on the decreased availability of abortions for low-income women into trend analyses (Partin and Palloni, 1994).

Statistical modeling and estimation are essential to improve our understanding of the sociodemographic differentials in infant mortality. Thus, data which allow for sophisticated modeling and controls are crucial. Currently, some of the best data available for this purpose are the national linked birth and death certificate files and the special natality surveys (NCHS, 1993; NCHS, 1985; Sanderson et al., 1991; Overpeck et al., 1992).

How Can Improved Indicators be Produced Over the Next Decade?

While an impressive system for the collection of infant mortality data is currently in place, several areas of improvement have been noted. Suggestions for areas in need of further study and refinement are presented below.

First, although the vital statistics system plays a valuable and indispensable role in the production of infant mortality statistics, further study is needed to evaluate the reporting completeness of this system. Out-of-hospital births and deaths and the nonuniform application of definitions of live births and fetal death could contribute to the underreporting of infant deaths (Kleinman, 1986). Additional research is needed to better evaluate the reporting problems and the degree of reporting completeness in the death registration system.

Second, cause of death information on death certificates should be improved. The following interventions have been recommended: 1) increased training opportunities and education regarding death certificate completion, including training in how to use the *International Classification of Diseases* and WHO selection rules for physicians (Carter, 1985; Rosenberg, 1989; 2) querying of physicians regarding questionable or suspect cases, which provides ongoing education and feedback and in turn improves the quality of cause of death information (Hopkins et al., 1989); and 3) initiation of a two-part death certificate; the first part would include only demographic information and would provide a quick way to register the death, while the second part would include an investigation form to be completed at a later date by qualified certifiers (Salmi et al., 1990). In addition, Carver et al. (1993) recommend that WHO selection rules be modified to allow short gestation priority over immediate causes of infant death.

Third, the quality and consistency of infant race/ethnicity coding on birth and death certificates needs to be improved. Several of the suggestions mentioned above (increased training, initiation of a two-part death certificate) may also improve the quality of race, ethnicity and other sociodemographic information on the death certificate. Another way to improve the quality of data on race/ethnicity would be to rely on maternal reports rather than the observation of the physician or other health professional.

Fourth, faster turn-around time is needed for information from the linked birth/death certificate files. The increasing use of the electronic transfer of data may assist in this process. Fifth, in addition, continued opportunities are needed to augment vital registration data by linking it with medical records information and/or survey questionnaire data. For instance, detailed information on maternal socioeconomic status that can be linked with vital records information is needed to further investigate the complex interaction of race and social class in regard to infant mortality.

Finally, research also is needed to identify data collection and definitional issues that might be related to cross-national differences in infant mortality. Additional data are needed to determine what portion of the observed higher level of infant mortality in the United States is due to procedural or methodological differences in the way the indicator is produced (Howell and Blondell, 1994).

B. Measures of Low Birth Weight

How is Low Birth Weight Currently Measured?

Definitions: While low birth weight is most commonly represented in terms of the proportion of infants born at weights less than 2,500 grams (or approximately 5.5 pounds), measures distinguishing very low birth weight infants (less than 1,500 grams) from moderately low birth weight infants (1,500-2,499 grams), and low birth weight due to intrauterine growth retardation from low birth weight due to premature delivery (before 37 weeks gestation) are also widely employed. The low birth weight "rate" refers to the percent of live births delivered at weights less than 2,500 grams in a given time period (usually a year). Measures of low birth weight distinguishing between premature and growth retarded infants have been variously defined. The most common definition involves a simple trichotomy where infants born both before 37 weeks gestation and at weights less than 2,500 grams (labeled "premature low birth weight"), and infants born at or after 37 weeks gestation and at weights less than 2,500 grams (labeled "intrauterine growth retarded low birth weight") are distinguished from infants of normal weight and gestation. Some health professionals and researchers have focused on more detailed definitions of the joint distribution of gestational age and birth weight. These more detailed definitions are frequently based on published fetal growth curves, such as those produced from an early study of births delivered in a Colorado hospital (Lubchenco et al., 1966), which summarize variation in the birth weights of infants delivered at various gestational ages. This information from the distribution of birth weights among infants delivered at various ages is commonly used to categorize infants into percentiles of birth weight for gestational age. More recently it has been argued, however, that an infant's birth weight should be expressed not in terms of divergence from some absolute standard, but rather in terms of standard deviations from population specific mean birth weights for gestational age (Wilcox and Russell, 1990).

The individual investigator's decision regarding which measure of low birth weight to employ will be guided not only by the level of detail desired, but also by the quantity and quality of the data available. For instance, investigators with only sparse data available will want to rely on less-detailed definitions. For reasons delineated below, definitions of low birth weight conditioned on gestational age should not be used unless resources are available to meticulously clean gestational age data for missing and implausible values.

Data Production: The primary source of data on low birth weight is the birth certificate. Data on registered births are published annually by the National Center Statistics in Vital Statistics of the United States, and in the Monthly Vital Statistics series under the title of "Advance Report of Final Natality Statistics" (see for instance National Center for Health Statistics, 1991). The latter summaries provide various information on birth weight including: 1) the proportion of births delivered in 500 gram birth weight categories by race and age of the mother; and 2) the proportion of infants born at very low, low, and high (4,000 grams or more) weights by maternal race and Hispanic origin. Since the 1989 revisions in the birth certificate, NCHS has published a new report entitled "Advance Report of Maternal and Infant Health Data." Along with a variety of other useful information on maternal and infant health, this report includes tabulations of: 1) the percent low birth weight by smoking status, age, and race of the mother; and 2) the percent low birth weight by maternal weight gain during pregnancy, period of gestation, and race of mother. National trends in low birth weight incidence can easily be examined with the use of the *Health, United States* series, which publishes national level data on the incidence, prevalence and distribution of a variety of health related behaviors and outcomes over time, and is also updated annually by NCHS (see for instance, NCHS, 1993). In addition to the above sources of national level data, most states provide county-specific information on the annual proportion of births delivered at weights less than 2,500 grams.

While vital records data provide the only continuous source of data on low birth weight, periodic sources of data include: 1) medical records matched with samples of births from local hospitals; 2) maternal reports of pregnancy histories obtained from social surveys such as the National Survey of Family Growth and the National Longitudinal Survey of Youth; 3) the National Natality Surveys; and 4) the 1988 National Maternal and Infant Health Survey. The advantages of the periodic sources of data mentioned above over vital registration data are that: 1) they enable the examination of low birth weight by subgroups such as family income and poverty status that cannot be identified in published vital statistics data; and 2) since individual records can be identified, they enable detailed analyses of the distribution and determinants of low birth weight not possible with aggregate level data. Furthermore, prior to 1989, these periodic surveys were the only source of data enabling the examination of the incidence of low birth weight by maternal health and health-related behaviors during pregnancy.

Data Quality: Low birth weight is generally used as an indicator of infant frailty. The validity of this measure as an indicator of frailty is well documented. Low birth weight is one of the strongest determinants of infant mortality (Shah and Abbey, 1971; Susser et al., 1972; Shapiro et al., 1980; Eberstein, 1984; McCormick, 1985; Tompkins, 1985; Rogers, 1989; Cramer, 1987; Haaga, 1989; Eberstein et al., 1990). Indeed, the 2,500 gram cutoff for low birth weight is the conventional comparison precisely because studies of birth weight-specific mortality have demonstrated that infant mortality rates rise sharply below this weight (Kramer, 1987; Puffer and Serrano, 1973; Saugstad, 1981). In addition to being a strong predictor of infant mortality, low birth weight is also associated with greater morbidity in the first few years of life (Shapiro et al., 1980; Hackman et al., 1983; Hayes, 1987), with certain neurological and developmental handicaps (Hayes, 1987; Fitzhardinge, 1976; Stewart et al., 1978; Hack et al., 1972; Papile et al., 1983; Ramey et al., 1978; Fitzhardinge and Steven, 1972; Harvey et al., 1982; Westwood et al., 1983), and with cognitive capacity, adaptive skills, and scholastic performance (McCormick et al., 1992; Baker et al., 1989; McCormick, 1985).

The quality of low birth weight data varies according to source. While certainly not error free (Horwitz and Yu, 1984; Romm and Putnam, 1981; Demlo et al., 1978; Hendrickson and Myers 1973), medical records are generally considered the gold standard for data of clinical importance. Examinations of the quality of data on low birth weight therefore frequently focus on comparing data obtained from the birth certificate with data recorded on medical records, using either the proportion of cases agreeing on both sources or the sample kappa statistic as the measure of reliability. In general, these studies suggest that the data obtained from birth certificates is of quite respectable quality. For instance, studies comparing birth weight data obtained from the birth certificate with birth weight data obtained from medical records report levels of agreement between the two sources ranging from 87 to 100 percent (Buescher et al., 1993; Querec, 1980; Piper et al., 1993; NCHS, 1985).

Studies examining the validity of maternal reports of birth weight have focused on comparing these reports with data obtained from the birth certificate. As with the studies comparing birth certificate and medical record data on birth weight, these studies suggest the quality of the birth weight data obtained from maternal reports is quite high, reporting levels of correspondence between the two sources ranging from 70 to 96% (Gayle et al., 1988; Tilley et al., 1985). Most studies interested in maternal and child health outcomes seek maternal report data within nine months of delivery, but some studies rely on data recalled potentially years after the birth of a child. Investigators relying on maternal report of low birth weight should be aware that the quality of maternal recall data generally tends to deteriorate over time (Oates and Forrest, 1984).

While the above discussion suggests that the available data on birth weight is of relatively high quality, it is not without its limitations. Researchers have identified several shortcomings of birth weight data. The first shortcoming involves the suspected under reporting of live births delivered at extremely low (i.e., less than 500 grams) weights. While it is presumed that the United States more completely reports very low birth weight infants than other countries (Howell and Blondel, 1994), some misreporting of these extremely low birth weight infants as stillbirths likely still occurs in our vital registration system. For this reason, infants weighing less than 500 grams are frequently excluded from analyses. The second shortcoming of available birth weight data involves the selective accuracy of data obtained from the birth certificate and maternal reports. Gayle et al. (1988) found that lower accuracy of birth weight reporting was associated with high risk profiles (low birth weight, preterm delivery, low APGAR scores, multiparity, low maternal education, black race, unmarried marital status, and young maternal age). This non-random accuracy of birth weight data may bias the results of analyses comparing birth weight across various population subgroups. The third shortcoming of birth weight data involves the common response bias of digit preference. David (1980) found in his analysis of 1975-1977 North Carolina birth certificates that the distribution of recorded birth weights demonstrated heaping at every quarter pound. This heaping should be taken into consideration when researchers are grouping birth weight into categories. To minimize biases, cut points should be made such that the range of birth weight in each category is centered around peaks in reporting. The final shortcoming of birth weight data involves errors occurring while the data from the birth certificate is key entered into computerized records. Brunskill (1990) identified three common types of errors in the coding of birth weight during key entry: 1) confusion of ounces for pounds; 2) mistaken reading of 1 pound as 11 pounds; and 3) errors in the placement of the decimal. All three of these reporting errors lead to systematic overreporting of extremely high birth weight infants and underreporting of very low birth weight infants.

The results from studies estimating the quality of birth weight data obtained from various sources suggest that while the simple dichotomous measure of low birth weight which distinguishes between infants delivered at weights less than 2,500 grams from those delivered at weights of 2,500 grams or more may be less precise than more detailed definitions, it does produce higher levels of correspondence in responses across various data collection instruments. The results of Gayle et al. (1988) mentioned earlier suggest that individuals employing data from either birth certificates or maternal reports collected within nine months of delivery can rely on at least 98% of cases being accurately categorized into low birth weight and normal birth weight categories. If definitions of birth weight conditioned on gestational age are preferred, however, data quality may be seriously compromised. The most common estimate of gestational age employed--the number of weeks between the date of delivery and the date of the last menstrual period (LMP) reported by the mother--suffers from some serious limitations. While this measure is the estimate provided on the standard birth certificate, it frequently produces high proportions of missing or incomplete information (Alexander and Cornely, 1987; David, 1980), and tends to display low levels of accuracy at the extremes of the gestational age distribution (Kramer et al., 1988; David, 1980). While ultrasound images are believed to provide more accurate estimates of gestational age than LMP data, the former estimate cannot be accurately ascertained for mothers receiving no prenatal care, and is considered inaccurate for mothers receiving their first prenatal care visit in the third trimester of pregnancy. Since not all pregnant women receive prenatal care before the third trimester, and some never receive any prenatal care (NCHS, 1993), sole reliance on ultrasound estimates of gestational age can lead to selective missing information and consequently biased results. The obstetrician's best estimate (OBE) of gestational age, which is based on both ultrasound images and LMP estimates when both are available, and LMP only when ultrasound images are not available, provides an attractive alternative to the sole reliance on either LMP or ultrasound measures. Inclusion of this estimate of gestational age on the birth certificate would likely improve the coverage and quality of available gestational age information.

Methods of Data Collection: The various methods of collecting data on birth weight afford different advantages and disadvantages. An important advantage of birth certificate data on birth weight is that it is collected continuously and disseminated annually, and therefore allows the examination of trends in incidence of low birth weight over time. Since low birth weight is a relatively rare event in the United States (approximately

7% of infants born in 1991 were delivered at low weights), birth certificate data also afford the advantage of providing enough cases to analyze differences in the distribution and determinants of low birth weight across various subgroups of the population. An important disadvantage of birth weight data obtained from both birth certificates and medical records is potential errors in the classification of mothers and infants into racial and ethnic subgroups. Since the information on maternal and infant race and ethnicity recorded on both the birth certificate and medical record may be determined by the observation of a physician or other health professional rather than maternal report, these data are likely a less valid measure of these characteristics than measures obtained from maternal reports. Other problems associated with race/ethnicity coding on vital records were discussed above. A disadvantage in birth weight data shared by all three sources (birth certificates, medical records, and maternal reports) is the questionable quality of gestational age data. As mentioned previously, if definitions of birth weight conditioned on gestational age are preferred, data quality may seriously be compromised.

How Should Low Birth Weight Indicators Be Produced?

Low birth weight is an incidence measure produced at the level of the individual, or infant. While researchers occasionally have measured the incidence of low birth weight at the level of the mother for the purposes of identifying women at high risk of delivering low birth weight infants in the future, this measure has only very specialized applications and has little utility in aggregate form. Aggregate figures of the incidence of low birth weight are currently summarized annually by both state and National Centers for Health Statistics for various subpopulations as described above. The availability of such continuous data on the incidence of low birth weight across various subgroups of the population is essential to the careful surveillance of infant health, and the monitoring of progress of national and local groups toward reaching Year 2000 objectives for infant health. The currently available disaggregations of low birth weight by severity (very low and moderately low) are particularly important. Research on low birth weight suggests that the determinants and consequences of these categories of birth weight differ, and that improvements in very low birth weight have lagged far behind improvements in moderately low birth weight over time (Kleinman and Kessel, 1987). The distribution of low birth weight according to maternal age, race and health behaviors should also continue to be disseminated. The substantial black-white gap in low birth weight has persisted for decades, and has actually widened rather than narrowed in recent years (Partin and Palloni, 1994). Subgroup information on race is particularly essential for tracking progress toward the national Year 2000 low birth weight goals. To help narrow the race gap in low birth weight, the Year 2000 objectives for absolute declines in low birth weight for blacks are greater than those targeted for whites.

Careful adjustment of low birth weight indicators for trends in other factors such as changes in the distribution of maternal age, marital status and education at birth, or changes in maternal health-related behaviors requires subgroup data on low birth weight not currently available in aggregate form. However, investigators can use published data on the characteristics of live births over time, in combination with regression analyses of the effects of these characteristics on low birth weight, to estimate the fit between various demographic trends and changes in low birth weight. This approach was recently used, for instance, to demonstrate the sensitivity of low birth weight trends to changes in fetal death rates over time (Partin and Palloni, 1994).

While much progress toward understanding the correlates and determinants of low birth weight has been made in the last 30 years, efforts to reduce low birth weight in this country have fallen short of expectations. If progress is to be made in reducing low birth weight, research on patterns and determinants of low birth weight must continue. The continued timely creation and availability of rich, nationally representative natality surveys such as those produced by the NCHS most recently in 1980 and 1988 is essential to this endeavor.

How Can Improved Indicators Be Produced Over The Next Decade?

While detailed information on the incidence and distribution of low birth weight is readily available at both the state and national level, and has been shown to be of particularly high quality, the above discussion suggests the need for improvement in several areas. Areas needing improvement and suggestions for how to achieve higher quality data on low birth weight in the future are delineated below.

First, high proportions of missing data and low levels of accuracy are important limitations of available gestational age data. David (1980) has offered the following suggestions for improving the coverage and accuracy of these data. Efforts should focus on improving gestational age reporting performance in the hospitals that tend to produce the records with the most errors. This might be done by instructing hospitals not to edit their gestational age data (i.e., reporting gestations that do not fit the clinical pattern as unknown) Partial information on gestational age should also be salvaged. Presently, if only month and year of LMP appear on the birth certificate, completed weeks gestation generally is coded as unknown. This LMP data, while incomplete, could be useful, and is present in most cases lacking full LMP data (David, 1980). In addition, standardized data surveillance programs at the state level could improve the completeness and accuracy of the birth files by checking unrealistic values for keying errors and by providing feedback on a regular basis to reporting hospitals about their performance in providing raw data. Finally, keystroke errors (e.g. confusing 1 pound for 11 pounds and misplacements of decimal points) might be minimized if birth weight pounds and ounces were recorded on separate lines or in separate boxes on the birth certificate (Brunskill, 1990).

Second, the underreporting of very low birth weight infants continues to compromise the quality of birth weight data. The standardized surveillance programs at the state level suggested by David (1980) for improving the coverage and quality of gestational age data could also be used to promote more accurate and thorough documentation of very low birth weight infants.

Third, the non-random accuracy of low birth weight and gestational age data may bias comparisons across population subgroups. Providing feedback on a regular basis to reporting hospitals about their performance in providing complete raw data for various populations subgroups with typically high rates of missing data could help promote higher quality data.

C. Measures of Prenatal Care Utilization

How is Prenatal Care Currently Measured?

Definitions: Prenatal health care refers to pregnancy-related services provided between conception and delivery, and may include monitoring the health status of the mother and fetus; providing information to foster optimal maternal health, dietary habits, and hygiene; and providing appropriate psychological and social support. Because information on the timing of the first prenatal care visit and the total number of prenatal care visits received represent the aspects of prenatal care most readily available to investigators, prenatal care is most often defined as a function of one or both of these pieces of information. The most common definition of prenatal care employed by investigators that combines these two pieces of information is the Kessner index. This definition of prenatal care adjusts the number and timing of prenatal care visits to gestational age and groups mothers into categories of "inadequate", "intermediate" and "adequate" care according to recommendations from the American College of Obstetricians and Gynecologists (ACOG) (1974). Cases with missing information on any one of the items making up this index are assigned to the inadequate care category. Modifications of the treatment of missing values on gestational age have been explored by other researchers with some success. For instance, many researchers have dealt with the problem of missing information on the exact day of the last menstrual period by assigning the 15th day of the month to that value. Studies employing this procedure suggest it does not substantially bias the direction of results (Binkin et al., 1985; Alexander et al., 1985). While several researchers have criticized the Kessner index for its lack of detail (Alexander and Cornely, 1987; Kotelchuck, 1987), it continues to be the most widely used measure of prenatal care.

While considerable effort has been expended to arrive at valid measures of the both the timing and quantity of prenatal care obtained by mothers, little attention has been paid to the distribution and content of the prenatal care visits obtained. As pointed out by Alexander and Cornely (1987, p. 250), one disadvantage of the Kessner index described above is that "women who initiate their first visit early, who do not return for care until late in pregnancy, and who do so because of complications resulting in a flurry of visits prior to delivery, would be indistinguishable from women making the same number of visits in an orderly fashion." They suggest future research consider the spacing of prenatal care visits along with the timing of the first visits and total number of visits. Other investigators are pushing for measures which consider the adequacy of content, as well as timing and quantity, in measures of prenatal care (Petitti et al., 1991; Hansell, 1991; Kogan et al., 1994a; Kogan et al., 1994b). The individual investigator's decision as to which definition will best suit their purposes will be shaped not only be the degree of accuracy desired, but also by the quality and availability of data on the various aspects of prenatal care.

Recently, Kotelchuck introduced a new method for assessing the adequacy of prenatal care Kotelchuck, 1994a). This new index--called the Adequacy of Prenatal Care Utilization Index (or APNCU Index)--was developed in response to some of the limitations of the Kessner Index. This new index measures prenatal care on two distinct and independent dimensions: the adequacy of the initiation of prenatal care (broken down by month prenatal care began rather than trimester), and the adequacy of the amount of prenatal care received once care has begun (measured as a percent of the number of ACOG-recommended prenatal care visits received during the time under care). The two dimensions are combined into a single summary index with the following four categories: adequate plus, adequate, intermediate, and inadequate. Using data from the 1980 National Natality Survey, a comparison of the APNCU Index with the Kessner Index revealed that 28.5% of women received different ratings, with the majority receiving a poorer rating on the APNCU Index (Kotelchuck, 1994a). Kotelchuck asserts that previous estimates of prenatal care in the U.S. may have overestimated its level of adequacy (Kotelchuck, 1994a, 1994b). Wise (1994) describes Kotelchuck's index as introducing "several important technical improvements over its predecessors" , providing a picture of the potential impact of prenatal care over the entire pregnancy period. In addition, the ability of the APNCU Index to "more accurately and comprehensively assess prenatal care utilization should enhance our understanding of the association between prenatal care utilization and birth outcomes in the United States" (Kotelchuck, 1994b).

Data Production: As with most other indicators of prenatal and infant health, the primary source of data on prenatal care is the birth certificate. Information on the timing of the first prenatal care visit and the total number of visits obtained by mothers extracted from birth certificates appear in the NCHS' Monthly Vital Statistics Report, the annual Natality volumes of Vital Statistics of the United States, and in the *Health, United States* series. For instance, the "Advance Report of Final Natality Statistics" appearing annually in the Monthly Vital Statistics Report series tabulates: 1) the number of live births to mothers beginning prenatal care in the first trimester by race and Hispanic origin; 2) the number of live births to mothers receiving late (after the second trimester) or no prenatal care by race and Hispanic origin; 3) the number of live births by month prenatal care began and age and race of mothers; and 4) the number of live births by the month prenatal care began, the total number of prenatal care visits received, and race of the mother. Trends in the measures of prenatal care provided in the Monthly Vital Statistics series can be examined with reference to various issues of *Health, United States*. In 1993, this volume included tabulations of the proportion of live births to mothers receiving early (initiated in the first trimester) and late (initiated in the third trimester) or no prenatal care in 1970, 1975, and 1980-91. Additionally, most states provide county-specific information on the proportion of births to mothers receiving late or no prenatal care in their annual vital statistics summaries.

While the continuously recorded birth certificate information described above is often more readily available to investigators, various periodic sources of data on prenatal care may be preferred by investigators desiring more detailed subgroup information. Periodic sources of prenatal care data include the various national surveys mentioned previously. The richest subgroup data available on prenatal care measures comes from the 1972 and 1980 National Natality Surveys. These surveys combined information on the timing of the first prenatal care visit and the total number of prenatal care visits obtained from the birth certificate, medical records, and maternal reports, and are one of the few sources of maternal and infant health data which provide information on income and poverty status. An important advantage of these surveys is the ability to compare data on prenatal care obtained from various sources. Two important disadvantage of these surveys are the fact that they are now somewhat dated and that the information on poverty was collected only for married mothers.

The 1988 Maternal and Infant Health Survey is an important source of information on the content of early prenatal care visits and includes information on the poverty status of both married and unmarried mothers

(Sanderson et al, 1991). In addition to the National Natality Surveys, and the National Maternal and Infant Health Survey (NMIHS), information on the timing of prenatal care can also be obtained from the NSFG. As with the 1988 NMIHS, the National Survey of Family Growth (NSFG) collected information on poverty status from both married and unmarried mothers. A less frequently exploited source of information on prenatal care is the National Longitudinal Survey of Youth (NLSY). The longitudinal nature of this data set enables comparisons of care seeking patterns among women with different profiles not possible in other data sources. For instance, analyses examining the effect of complications in prior pregnancies to prenatal care sought in the current pregnancy are possible with this data source, and might be helpful in resolving remaining questions regarding the extent of selection into prenatal care described below.

Data Quality: Prenatal care presumably improves pregnancy outcomes by serving as a screening mechanism for high risk pregnancies. If measures to prevent poor outcomes are to be effectively invoked, high risk pregnancies must be identified early and monitored regularly. Prenatal care may also improve pregnancy outcomes by modifying certain maternal behaviors observed (such as smoking, drinking, and poor nutritional habits) that may threaten the healthy development of the fetus. Studies investigating the association between prenatal care and pregnancy outcomes generally show that mothers lacking prenatal care are more likely to deliver low birth weight infants (Eisner et al., 1979; Greenberg 1983; Leveno et al., 1985) and experience infant death (Leveno et al., 1985) than mothers who have had at least some prenatal care. However, while existing studies suggest that prenatal care is an important determinant of both prenatal and infant health, the validity of the prenatal care measures most commonly employed may be severely limited. Since there have not been any carefully conducted clinical trials of the efficacy of prenatal care, investigators have had to test the extent to which prenatal care actually represents an indicator of favorable health inputs with the data available (primarily that from the birth certificate). When randomized clinical trials are not feasible (as is the case with prenatal care, which is so much a part of common obstetric practice that it cannot ethically be withheld from mothers), a rigorous test requires careful standardization and sophisticated modeling, as described below.

While aggregate measures of the timing and quantity of prenatal care visits are readily available, using these measures in the absence of any adjustments may leave investigators with invalid measures. Arriving at valid measures of prenatal care requires surmounting several methodological challenges. The first challenge stems from the fact that the number of prenatal care visits obtained by the mother is restricted by the length of gestation. This simultaneity of prenatal care and gestational age (often referred to as the "preterm bias" effect) makes it difficult for the investigator to distinguish whether the length of gestation was cut short because the number of prenatal care visits was inadequate, or whether the number of prenatal visits was cut short as a result of a short gestation caused by other factors. One way to disentangle these associations is to define prenatal care as a function of the length of gestation. This is the approach used to create the various index measures discussed above. Another solution involves using the predicted number of prenatal care visits expected by a given gestational age (e.g., 37 weeks), which are estimated from a model for prenatal care, in the model assessing the effects of prenatal care on pregnancy outcomes. This approach has been adopted by several economists (see for instance, Guilky et al., 1989).

The second challenge to arriving at a valid measure of prenatal care involves the fact that, because the amount of prenatal care obtained represents at least in part behavioral choices of the mother, any observed association between prenatal care and health outcomes may be due partially, if not entirely, to self selection. The most common strategy employed for correcting for the selective nature of prenatal care is the instrumental variable approach. This approach corrects for the selective nature of prenatal care by regressing the number of prenatal care visits (or some other indicator of prenatal care) on various exogenous factors which serve as instruments for identifying the unobserved characteristics of the mother which are both related to the pregnancy outcome and to the amount of prenatal care the mother seeks. The success of this approach in adjusting for the biases introduced by these unobserved factors is of course contingent upon obtaining a suitable array of instruments. One requirement is that the equation predicting prenatal care include an assortment of exogenous factors which are associated with the amount of prenatal care received but not with the outcome of the pregnancy itself. Investigators have generally relied on information describing the availability of care, such as number of prenatal care clinics in the area and distance to the closest clinic, to satisfy this requirement. While those studies actually estimating the biases introduced by the selective nature of prenatal care provide strong evidence for the presence of adverse selection into prenatal care (Rosenzweig and Schultz, 1983; Guilky et al., 1989; Joyce, 1990; Grossman and Joyce, 1990), the potential for positive selection in other studies cannot be ignored.

In addition to the formidable obstacles to obtaining valid measures of prenatal care discussed above, a number of other factors threaten the overall quality of available prenatal care measures. The most serious threat to the quality of prenatal care data is the high proportion of cases with missing information on prenatal care and gestational age (Alexander et al., 1991; Forrest and Singh, 1987; Piper et al., 1993). The quality of prenatal care data is also called into question by studies finding low levels of correspondence in prenatal care information obtained from different sources (Buescher et al., 1993; Querec, 1980; Forest and Singh, 1987; Piper et al., 1993). A final threat to the quality of the most commonly employed measures of prenatal care are the limitations of gestational age data delineated above. While efforts to control for the preterm bias effect are a necessary step in arriving at valid measures of prenatal care, investigators should carefully inspect gestational age information for non-random patterns of missing data and implausible values.

Methods of Data Collection: No source or definition of prenatal care is flawless. For instance, while measures of prenatal care conditioned on gestational age and corrected for potential self selection are thought to be more valid than uncorrected measures, they make much greater demands on data typically limited by high proportions of missing data and low levels of quality. Similarly, while medical records data are widely considered a more valid source of information on the timing, content, and quantity of prenatal care received by mothers than birth certificates and maternal recall, the higher rate of missing information from this source (Forrest and Singh, 1987) dampens its overall advantage over others sources. The birth certificate may be the preferred source of information for investigators seeking a continuous source of temporal data, but suffers the disadvantage of providing less subgroup detail than most periodic sources of information. If the distribution and determinants of prenatal care, or the association between prenatal care and health outcomes is of primary interest, investigators may want to turn to more detailed sources of data such as the National Natality Surveys, the NMIHS, the NSFG or the NLSY. Although the Natality Surveys provide the richest data available on prenatal care, combining reports from birth certificates with those from medical records and maternal interviews, the most recent wave (1980) is now quite dated. The 1988 National Maternal and Infant Health Survey provides more timely data, but hospital record section of the survey containing detail on the distribution and content of prenatal care visits has not been released. Investigators interested in a prospective source of information can turn to the NLSY, but will have to rely on maternal reports of somewhat limited prenatal care data.

How Should Prenatal Care Indicators Be Produced?

State and national level data on prenatal care are generally summarized by prevalence measures of the proportion of women delivering in a given year that received prenatal care during their pregnancy. Information on whether care was received and the trimester in which the first visit was made should continue to be made available by maternal age and race. However, given the dependence of prenatal care receipt on the gestational length of pregnancies, measures of prenatal care standardized by gestational age should also be provided in addition to the currently available unstandardized measures. Standardized measures will allow investigators to distinguish subgroup differences and temporal trends in prenatal care receipt from patterns due to gestational age. Since the proper measurement of prenatal care may require statistical modeling and adjustments beyond the scope of many investigators, the availability of these standardized measures will likely be invaluable to investigators lacking the resources to estimate standardized measures of prenatal care themselves.

How Can Improved Indicators be Produced Over the Next Decade?

Of the three priority indicators of prenatal and infant health discussed here, prenatal care represents the indicator with the greatest overall need for improvement. The shortcomings of existing prenatal care data delineated in the above discussion reflect a lack of knowledge about prenatal care. In order to determine where to focus our efforts for arriving at improved indicators of prenatal care, we need to strengthen our understanding of the association between prenatal care and favorable health outcomes. In particular, we need to examine the content of prenatal care closely to identify those components most essential to ensuring healthy outcomes for both the mother and fetus (Nagey, 1989). We also need to examine the distribution of prenatal care visits across gestational age to determine the patterns most likely to ensure healthy outcomes. Finally, further research must also be conducted to better understand the potential for selection effects in specific populations.

In addition to further research to strengthen our understanding of prenatal care, the following suggestions should help us achieve better indicators of prenatal care in the future. First, what a prenatal visit actually represents needs to be defined more clearly. This will help to achieve a more reliable and valid measure of prenatal care. For example, it is currently unclear whether a visit for a pregnancy test should be considered as the first prenatal care visit.

Second, missing information on prenatal care and gestational age need to be reduced. Doing so in all data sources will greatly enhance the quality of data on prenatal care. The current amount of missing data on these factors has a great impact on the distribution of these variables. Third, improvements in the overall quality of gestational age data are desperately needed. As mentioned previously, inclusion of the obstetrician's best estimate of gestational age on the birth certificate would likely greatly improve the quality of available gestational age information. Reductions in the amount of missing information in gestational age data can be achieved via the guidelines recommended by David (1980) and mentioned above. In addition, standards for cleaning gestational age data published by NCHS for implausible values would be invaluable.

Finally, in addition to the unstandardized measures of prenatal care utilization that are currently available, NCHS should provide measures of prenatal care standardized for gestational age, with the resulting standardization protocols made available for others to emulate.

IV. ADDITIONAL PRENATAL AND INFANT HEALTH INDICATORS

In addition to the three measures selected as the priority indicators, there are numerous other indicators that also are useful in assessments of prenatal and infant well-being (recall Table 1). A brief discussion of these other indicators follows, including issues related to data collection and monitoring.

Fetal mortality rates: Pregnancies do not always end in a live birth. In the United States, the term "fetal death" is used to define the events commonly referred to as miscarriage, induced abortion, and stillbirth (Shryock and Siegel, 1976). The fetal mortality rate is generally disaggregated into the fetal death rate (defined as the number of deaths of 20 weeks gestation or more per 1,000 live births and fetal deaths), the late fetal death rate (defined as the number of fetal deaths of 28 weeks gestation or more per 1,000 live births and fetal deaths), and the perinatal mortality rate (defined as the number of late fetal deaths and infant deaths under 7 days of age per 1,000 live births and late fetal deaths). Fetal deaths at 20 or more weeks of gestation are registered separately from other deaths, through the use of a fetal death report rather than a death certificate.

Fetal mortality is an important indicator of prenatal health and is associated with preventable or manageable aspects of maternal morbidity (Pritchard and MacDonald, 1980; Partin, 1993). Although Year 2000 maternal and infant health goals include reducing the overall and race-specific rates of fetal death, assessing progress towards these goals is challenged by limitations of available data. While the United States tends to have more complete reporting of perinatal deaths than other countries (Howell and Blondel, 1994), underreporting remains a serious limitation of all measures of fetal mortality in this country (Kleinman, 1988).

The majority of early fetal deaths (those occurring prior to 28 weeks gestation) are due to spontaneous and induced abortions, both of which are particularly subject to underreporting (Jones and Forrest, 1992; Wilcox and Horney, 1984). The early *spontaneous* abortions (or miscarriages) that get reported are selectively those that are associated with severe complications and/or are known to medical staff. A long-term study of menstrual cycles found that no more than 75% of prospectively recorded spontaneous abortions were later recalled by women, and that early abortions were recalled less often than later ones (Wilcox and Horney, 1984).

Jones and Forrest (1992) found that *induced* abortion reporting was highly deficient in the NSFG and the NLSY. While the Alan Guttmacher Institute compiles data from abortion providers, assessments of this data suggest that counts based on these data slightly underestimate the true national total (Henshaw and VanVort, 1990). The Centers for Disease Control and Prevention (CDC) also have an abortion surveillance program, through which data on induced abortion are compiled from 50 states, New York City, and Washington, D.C. Although the CDC program covers the entire country, the overall numbers reported by some states are incomplete and not all states report all characteristics.

NCHS publishes annual rates of fetal mortality, late fetal mortality, and perinatal mortality. Given the measurement problems associated with counting early fetal deaths, the quality of data regarding **late** fetal mortality is likely the highest. The rigor with which late fetal deaths are reported, however, tends to vary by hospital. In addition, definitions of fetal mortality are dependent upon an assessment of gestational age, which--as discussed above---is subject to its own set of deficiencies.

Measures of Maternal Health During Pregnancy: There are several ways in which maternal health (and thus the health of the fetus) during the course of a pregnancy can be measured. Examples of this type of indicator include: 1) rates of pregnancy complications such a gestational diabetes, preeclampsia, and placenta previa; 2) maternal weight gain during pregnancy as a measure of adequate nutrition; and 3) maternal tobacco, alcohol and other substance use and abuse during pregnancy. Each of these three measures are discussed briefly below.

Medical complications experienced by the mother during pregnancy contribute to maternal, fetal, and neonatal mortality, and are associated with preterm and low birth weight delivery (Burrow and Ferris; 1988; Placek, 1977; Hutchkins et al., 1984; Powell-Griner and Rogers, 1987). The 1989 Standard Certificate of Live Birth includes--in checkbox format--information on both common underlying health complications present during pregnancy and various complications of labor and delivery. Studies conducted in Washington state suggest that the new checkbox format on the 1989 birth certificate will improve the reporting of pregnancy, labor and delivery complications over the open-ended format used previously (Frost, 1984). In general, birth certificate information on maternal pregnancy complications is not as accurate as medical records data, but provides a more accurate picture than maternal self report. The advantage of using birth certificate information is that it is available for all births at the national, state and local level. Caution should be exercised when using these data from vital records, however, since the sensitivity of the data on maternal medical risk factors and complications of labor and delivery is generally low (Piper et al., 1993).

Since the early 1980's, physicians have been recommending that women gain between 22 and 27 pounds during pregnancy, primarily because low maternal weight gain (especially under 16 pounds) has been found to be associated with poor pregnancy outcomes (Taffel and Keppel, 1986). An adequate assessment of maternal weight gain needs to consider both the height of the mother and gestational age at delivery. Although maternal weight gain is reported on the birth certificate, the resulting information is most often based on self-reported information of questionable quality. Medical records may provide more objective documentation of maternal weight gain during the pregnancy. These data, however, are limited by the fact that information on maternal weight is present only for women receiving prenatal care and is missing for the time period before prenatal care starts.

The use and abuse of tobacco, alcohol and other drugs during pregnancy has been associated with a variety of negative pregnancy outcomes. Smoking during pregnancy is strongly associated with low birth weight and other negative birth outcomes (Malloy et al., 1988; Eisner et al, 1979). Heavy alcohol consumption is associated with fetal alcohol syndrome (Rosett and Weiner, 1984), and cocaine use is associated with fetal distress, impaired fetal growth, and may result in long-term developmental and behavioral problems during and after infancy (Howard et al., 1989; MacGregor et al., 1987; Zuckerman et al., 1989). Maternal substance use during pregnancy is collected via the birth certificate, is documented in most medical records, and can be gathered through maternal/child health survey research. The quality of data on maternal smoking, alcohol use and illicit drug use during pregnancy has been fairly well studied. In general, the quality of self-reported smoking behavior is considered to be high (Patrick et al., 1994). It is possible, however, that women with poor pregnancy outcomes are more motivated to falsely report their tobacco use than women with normal, healthy babies. In addition, research suggests that the form and content of self report questions regarding risk behaviors strongly influences the responses obtained (Petitti, Friedman and Kahn, 1981). Not surprisingly, the comparison of self report and urine assays suggests that neither method of assessment is highly sensitive for illicit drug use (e.g. marijuana and cocaine) during pregnancy (Frank, 1988). Combining biochemical information with self-report data yields the highest-quality data available on maternal substance use during the prenatal period.

APGAR Scores: The APGAR score (named after Dr. Virginia Apgar) is a method for assessing physiological signs that denote the condition of an infant during the first critical minutes of life (including heart rate, respiratory effort, reflex irritability, muscle tone, and color). The score is the most commonly-used tool for assessing the physical status of newborns, and is a reportable item on the Standard Certificate of Live Birth. Newborns are given an APGAR rating at one minute and again at five minutes after birth. In general, the APGAR score has high specificity (a healthy newborn will not have a low or poor score) but lacks sensitivity (infants at risk of mortality, neurologic defects, or metabolic acidosis may receive high or good scores) (Kessel et al., 1988; Myers and Gleicher, 1988; NCHS, 1989). Both the one minute and five minute scores are good predictors of mortality and neurologic abnormalities in infants with normal birth weight. These scores, however, are poor predictors of outcomes for low birth weight and other high risk babies (American College of Obstetricians and Gynecologists, 1985; American Society for Human Genetics, 1987). Despite these limitations, APGAR scoring results have proved to be useful in obstetrical practice research. For example, results from this scoring method were instrumental in discouraging the use of narcotic analgesia and sedation or general anesthesia for deliveries. Studies comparing APGAR scores from birth certificates with scores from the medical records found very high rates of comparability (Buescher, 1993; Piper et al., 1993).

Congenital Anomalies: Rates of congenital anomalies or birth defects provide a picture of the proportion of children that are born with malformations or specific health conditions. It is estimated that approximately 3 percent of all live births have one or more major malformation or defects (Hexter et al.). Cause-specific rates of anomalies show the distribution of different types of birth defects and can be used to assess the number of children born with serious malformations. The surveillance of birth defects, however, is challenged by limitations associated with the two major sources of data: birth certificates and hospital discharge diagnoses for newborns (Calle and Khoury, 1994; Frost et al., 1984; Minton and Seegmiller, 1986). Many anomalies, such as cleft lip and palate, hydrocephalus, congenital hip dislocation, and Down's syndrome, are easily identified at birth. Many other types of anomalies, however, are not immediately apparent during the first days of life, including certain central nervous system disorders, genitourinary disorders, and heart malformations. Estimates of birth defect rates using data collected at the time of birth severely underestimate the overall rate of defects and several specific types. (Hexter et al., 1990; Minton and Seegmiller, 1986). In an effort to improve surveillance statistics and analytic studies of birth defects, most states have implemented birth defects registries or monitoring systems which gather reports on congenital defects from the time of birth and beyond.

Measures of Infant Morbidity: The level of infant morbidity--or the degree to which infants experience illness and disease--can be measured in a plethora of ways. Measures such as the proportion of live births admitted or transferred to neonatal intensive care units give an indication of the proportion of newborns needing specialized and intense care. Incidence rates for various illnesses indicate the number of new cases of specific diseases during the first year of life. The reporting of several diseases (e.g. vaccine-preventable diseases such as diphtheria, pertussis and measles, and chronic conditions such as cancer) is required by law in most states, making incidence rates readily available. Information on infant diseases or conditions that are not reportable (such as gastroenteritis, respiratory diseases, or epilepsy) is more difficult to obtain. For example, respiratory syncytial virus (RSV) is considered the major lower respiratory tract pathogen of infancy and early childhood throughout the world, and is the major cause of bronchiolitis and pneumonia in young age groups (Chanock et al., 1984). Since RSV is a leading cause of hospitalization among infants in the United States, data on this source of morbidity can be obtained from hospital discharge data. Since not all cases of the disease are hospitalized, however, available data on overall incidence, length of illness and treatment costs are severely limited.

Rates of immunization during infancy indicate the degree to which infants are being protected from serious childhood diseases, and also indicate the proportion of babies that are coming in contact with health professionals during infancy. Finally, the abuse and neglect of children is a serious and disturbing problem in our society. Data on the numbers of infants found to be abused or neglected seriously underestimate the prevalence of these horrors, as the majority of cases escape detection by police, health care providers, or child welfare authorities.

Measures of Growth and Development: There are several methods by which infants' physical growth can be measured. Changes in height and weight over time can be assessed, from which measures of various growth problems such as stunting and wasting can be produced. Motor skills acquisition and cognitive development can also be measured in many ways. A problem with measures of growth and development during infancy, however, is that the distribution of what is considered normal is quite wide, making it difficult to define rates of growth and development that indicate a lack of health or well-being. Most infants crawl by their first birthday, and many are also walking. The inability to crawl by age one, however, is not necessarily a cause for alarm or concern. Similarly, most infants have several words that they use consistently by their first birthday. Not using words by age one, however, is not a sign of developmental delay. Pediatricians are reluctant to label infants whose rates of physical, motor or cognitive development are relatively slow as problematic, since variation in rates of development is so great. Therefore, to be meaningful, indicators of growth and development must focus on extremes.

By focusing on priority indicators of prenatal and infant health, our discussion of measures is admittedly incomplete. We make no explicit mention, for instance, of the various indirect indicators of prenatal and infant well being (i.e. those measures and markers that are linked to or associated with prenatal and infant health and are therefore indirectly related to the well being of very young children). Examples of indirect indicators include: unintended pregnancy rates, teen pregnancy rates, maternal mortality rates, breast feeding rates and practices, postpartum substance abuse rates, antenatal care issues (such as rates of screening for genetic disorders and other disabling conditions), and social issues (such as the proportion of pregnant women and infants living in poverty, and the proportion of pregnant women and infants without health insurance). Although we did not select any of these indirect indicators as being among the priority indicators, they are nonetheless very important to health and well-being during the prenatal and infant time periods.

V. SUMMARY AND CONCLUSIONS

There are numerous indicators available to those who wish to assess the status of prenatal and infant health in the United States. Both direct and indirect indicators of health status are available at the local, state and national level. Overall, the state of indicators for prenatal and infant health is impressive, reflecting the historic interest that government, public health (including researchers), child advocates, and the general public have had in using various statistical indicators as measures of the health and social well being of children in our society. In addition, the state of prenatal and infant health indicators reflects the advanced state of our national and state vital registration systems.

The state of the three priority indicators (measures of infant mortality, low birth weight, and prenatal care utilization) was discussed in great detail. Although these indicators are already widely used to assess levels, trends and patterns in prenatal and infant health, each indicator is currently experiencing its own set of problems related to data quality and data dissemination. Some of these problems are in need of further study and investigation before detailed recommendations for improvement can be made. Specifically, additional study is needed in the following areas: 1) the impact of the 1989 revised standard certificates for live births and deaths on data quality; 2) the state of completeness of birth and death certificate reporting; and 3) the association between prenatal care content and favorable health outcomes. Conversely, other problems have been studied to the extent that concrete recommendations for improvement have been made, and efforts to improve existing data can now be implemented. These include the following areas: 1) improvements in cause of death reporting on death certificates; 2) improvements in the coding of race and ethnicity on birth and death certificates; 3) improvements in the quality of birth weight and gestational age data; 4) reductions in the amount of missing data on prenatal care and gestational age on the birth certificate; and 5) the development and acceptance of a standardized measure of prenatal care

David (1980) argues that some of the shortcomings in vital registration data on prenatal and infant health will not significantly improve until the dramatic differences in prenatal care use between mothers with accurate and inaccurate data have been addressed. For this to be accomplished, health professionals must develop a greater proficiency at bridging the communication barriers that separate them from socioeconomically disadvantaged mothers in their patient populations. This is a formidable yet crucial challenge.

In the United States, there is a vast array of resources available to those who want to document and/or investigate the determinants of prenatal/infant health and illness. National vital records data and national surveys (such as the National Natality Surveys, National Health Interview Survey, and National Survey of Family Growth) provide researchers with a plethora of opportunities for investigation. The third wave of the National Health and Nutrition Examination Survey, with its oversampling of children under the age of 36 months, will provide new and unique data on the health status of infants on a national level. Increased research activity could also be realized at the state and local level, taking advantage of vital statistics and medical records data that are currently available. In addition, however, we believe that some of the most important research that needs to be done is that which will offer instruction on how to move beyond the mere production of statistical indicators and measures. Specifically, research is needed to guide the process of transforming indicators and descriptive research findings into policy recommendations and evaluation tools. Only then will the assessment and production of prenatal and infant health indicators result in the larger goal of actually improving the health and well-being of children in our society.

Table 1: List of Direct Indicators of Prenatal and Infant Health

Key Direct Indicators

Measures of Infant Mortality

Measures of Low Birth Weight

Measures of Prenatal Care Utilization

Other Direct Indicators of Prenatal Health

Fetal Mortality Rates

Measures of Maternal Health During Pregnancy:
    Rate of maternal pregnancy complications
    Maternal weight gain during pregnancy
    Maternal tobacco, alcohol and other drug use during pregnancy

APGAR Scores

Congenital Anomaly Rates

Measures of Infant Morbidity:
    Proportion of infants admitted to Neonatal Intensive Care Unit
    Incidence rates of illnesses during infancy
    Immunization rates
    Incidence rates of infant abuse and neglect

Measures of Growth and Development:
    Measures of physical growth (height/weight) during infancy
    Measures of motor skills acquisition during infancy
    Measures cognitive development during infancy

Table 2: Sources of Data for Indicators of Prenatal and Infant Health

Type of Data	Major Data Sources
Vital Registration Data	State and local data from birth certificates, death certificates, and fetal death reports
	National data from National Vital Statistics Program
	National linked files of live births and infant deaths
Medical Records Data	Local patient medical charts
	Local patient laboratory and procedure records
	Local patient billing records
	State hospital discharge data
	National Hospital Discharge Survey
Social Survey Data	State and local surveys
	National Survey of Family Growth
	National Longitudinal Survey of Youth
	National Health Interview Survey
	National Health and Nutrition Examination Survey
Combined Data Sources	National Natality Surveys
	National Maternal and Infant Health Survey

REFERENCES

Alexander GR and Cornely DA. Prenatal care utilization: its measurement and relationship to pregnancy outcome. *Am J Prev Med*. 1987;3:243-53.

Alexander GR, Tompkins ME, Altekruse JM and Hornung CA. Racial differentials in the relation of birth weight and gestational age to neonatal mortality. *Public Health Rep*. 1985;100:539-47.

Alexander GR, Tompkins ME, Cornely DA. Gestational age reporting and preterm delivery. *Pub Health Rep*. 1990;105:267-275.

Alexander GR, Tompkins ME, Peterson DJ and Weiss J. Source of bias in prenatal care utilization indices: implications for evaluating the Medicaid expansion. *Am J Public Health*. 1991;81:1013-1016.

American College of Obstetricians and Gynecologists. *Manual of Standards in Obstetic-Gynecological Practice*. Chicago: the College, 1974.

American College of Obstetricians and Gynecologists, Department of Professional Liability. *DPL Alert: Professional Liability Implications of AFP Tests*. Washington DC: the College, 1985.

American Society of Human Genetics. Policy statement for maternal serum alpha-fetoprotein screening programs and quality control for laboratories performing maternal serum and amniotic fluid alpha fetoprotein assays. *Am J Human Genetics*. 1987;40:75-82.

Baker DJP, Osmond C, Winter PD, Margetts B and Simmond SJ. Weight in infancy and death from ischaemic heart disease. *Lancet*. 1989;9:578-580.

Becerra JE, Hogue CJR, Atrash HK, Perez N. Infant mortality among Hispanics: a portrait of heterogeneity. *JAMA*. 1991;265:217-221.

Binkin NJ, Williams RL, Hogue CTR and Chen PM. Reducing black neonatal mortality: will improvement in birth weight be enough? *JAMA*. 1985;253:372-5.

Brunskill AJ. Some sources of error in the coding of birth weight. *Am J Public Health*. 1990;80:72-73.

Buetow SA. The perinatal autopsy: its conduct and reporting in Australia. *Med J Aust*. 1992; 156:492-494.

Buescher PA, Taylor KP, Davis MH and Bowling JM. The quality of the new birth certificate data: a validation study in North Carolina. *Am J Public Health*. 1993;83:1163-1165.

Burrow GN and Ferris TF. *Medical Complications During Pregnancy, 3rd Edition*. WB Saunders Co., Philadelphia, 1988.

Calle EE, Khoury MJ. Completeness of the discharge diagnoses as a measure of birth defects recorded in the hospital birth record. *Am J Epidemiol*. 1994;134:68-77.

Carnegie Task Force on Meeting the Needs of Young Children. *Starting points: meeting the needs of our youngest children: the report of the Carnegie Task Force on Meeting the Needs of Young Children*. New York: Carnegie Corporation of New York, 1994.

Carter JR. The problematic death certificate. *N Engl J Med*. 1985;313:1285-1286.

Carver JD, McDermott RJ, Jacobson NH, Sherin KM, Kanarek K, et al. Infant mortality statistics do not adequately reflect the impact of short gestation. *Pediatrics*. 1993;92:229-232.

Comstock GW, Markush RE. Further comments on problems in death certification. *Am J Epidemiol*. 1986;124:180-181.

Chanock RM, Hyun WK, Brandat CD, Parrott RH. Respiratory syncytial virus. In Evans AS, editor, *Viral Infections of Humans: Epidemiology and Control*. New York: Plenum Medical Book Company, 1984.

Children's Defense Fund. *The State of America's Children 1992*. Washington, D.C.: Children's Defense Fund, 1992.

Children's Defense Fund. *The State of America's Children Yearbook 1994*. Washington, D.C.: Children's Defense Fund, 1994.

Cramer J. Social factors and infant mortality: identifying high risk groups and proximate causes. *Demography*. 1987;24:299-322.

David RJ. Did low birth weight among U.S. blacks really increase? *Am J Public Health*. 1986:76:380-384.

David RJ. The quality and completeness of birth weight and gestational age data in computerized birth files. *Am J Public Health*. 1980;70:964-973.

Demlo LK, Campbell PM, and Brown SS. Reliability of information abstracted from patient's medical records. *Medical Care*. 1978;16:995-1005.

Eberstein I, Nam C, and Hummer R. Infant mortality by cause of death: main and interaction effects. *Demography*. 1990;27:413-430.

Eberstein I. Racial differences in infant mortality by cause of death: the impact of birth weight and maternal age. *Demography*. 1984;21:309-317.

Eisner V, Brazie JV, Pratt MW, and Hexter AC. The risk of low birth weight. *Am J Public Health*. 1979;69:887-893.

Fitzhardinge PM and Steven EM. The small-for-date infant. II. Neurological and intellectual sequelae. *Pediatrics*. 1972;50:50.

Fitzhardinge PM. Follow-up studies of low birth weight infants. *Clinical Perinatology*. 1976;3:503.

Forrest JD and Singh S. Timing of prenatal care in the United States: how accurate are our measurements? *Health Serv Res*. 1987;22:235-253.

Frank DA et al. Cocaine use during pregnancy: prevalence and correlates. *Pediatrics*. 1988;82:888-895.

Freedman MA, Gay GA, Brockert JE, Potrzebowski PW, Rothwell CJ. The 1989 revisions of the US standard certificates of live birth and death and the US standard report of fetal death. *Am J Public Health*. 1988;78:168-172.

Frost F, Starzyk P, George S, McLaughlin JF. Birth complication reporting: the effect of birth certificate design. *Am J Public Health*. 1984;74:505-506.

Frost F, Jennings T, Starzyk P. Completeness of infant death registration for very low birth weight infants: Washington state 1978-79. *Am J Public Health*. 1982;740-741

Gable CB. Reviews and commentary: a compendium of public health data sources. *Am J Epidemiol*. 1990;131:381-394.

Gayle HD, Yip R, Frank MJ, Nieburg P, and Binkin NJ. Validation of maternally reported birth weights among 46,637 Tennessee WIC program participants. *Pub Health Rep*. 1988;103:143-147.

Geronimus AT, Anderson HF and Bound J. Differences in hypertension prevalence among US black and white women of chidbearing age. *Pub Health Rep*. 1991;106:393-399.

Greenberg RS. The impact of prenatal care in different social groups. *Am J Obstet Gynecol*. 1983;145:797-801.

Grossman M and Joyce TJ. Unobservables, pregnancy resolutions, and birth weight production functions in New York City. *J Politic Econ*. 1990;985-983.

Guilky DK, Popkin BM, Akin JS and Wong EL. Prenatal care and pregnancy outcome in Cebu, Phillipines. *J Devel Economics*. 1989;30:241-272.

Haaga J. How are maternal age, parity, and birthspacing related to infant health? A review of biomedical evidence. Background paper prepared for the National Academy of Sciences, Committee on Population, Working Group on the Health Consequences of Contraceptive Use and Controlled Fertility, 1989.

Hack M, Fanaroff AA, and Merkatz IR. The low birthweight infant -evolution of a changing outlook. *N Engl J Med*. 1983;301:1162.

Hackman E et al. Maternal birth weight and subsequent pregnancy outcome. *JAMA*. 1983;250:2016.

Hahn RA, Mulinare J, Teutsch SM. Inconsistencies in coding of race and ethnicity between birth and death in U.S. infants. *JAMA*. 1992;267:259-263.

Hansell MJ. Sociodemographic factors and the quality of prenatal care. *Am J Public Health*. 1991;81:1023-1028.

Harlow SD, Linet MS. Reviews and commentary: agreement between questionnaire data and medical records: the evidence for accuracy of recall. *Am J Epidemiol*. 1989;129:233-248.

Harvey D, Prince J, Bunton J, Parkinson C, and Campbell S. Abilities of children who were small-for-gestational age babies. *Pediatrics*. 1982;69:296.

Haub C, Yanagishita M. Infant mortality: Who's number one? *Population Today*. 1991;March:6-9.

Hayes CD (ed.) Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing, vol. 2, Working Papers, National Academy of the Sciences Panel on Adolescent Pregnancy and Childbearing. Washington D.C., National Academy Press, 1987.

Healthy People 2000: National Health Promotion and Disease Prevention Objectives. Washington, D.C.: U.S. Government Printing Office; 1990. U.S. Department of Health And Human Services. Publication number 91-50212.

Hendrickson I and Myers J. Some sources and potential consequences of errors in medical data recording. *Meth Inform Med*. 1973;12:38-45.

Henshaw SK and VanVort J. Abortion services in the United States, 1987 and 1988. *Fam Plan Perspec*. 1990;22:102-42.

Hewson D, Bennett A. Childbirth research data: medical records or women's reports? *Am J Epidemiol*. 1987;125:484-491.

Hexter AC, Harris JA, Roeper P, Croen LA, Krueger P, Gant D. Evaluation of the hospital discharge diagnoses index and the birth certificate as sources of information on birth defect. *Pub Health Rep*. 1990;105:296-307.

Hopkins DD, Grant-Worley JA, Bollinger TL. Survey of cause-of-death query criteria used by state vital statistics programs in the US and the efficacy of the criteria used by the Oregon vital statistics program. *Am J Public Health*. 1989;79:570-574.

Horwitz RI and Yu EC. Assessing the reliability of epidemiologic data obtained from medical records. *J Chronic Dis*. 1984;37:825-831.

Howard J, Beckwith L, Rodning C and Kropenske V. National Center for Clinical Infant Programs. The development of young children of substance abusing parents: Insights from seven years of intervention and research. *Zero to Three*. 1989;4:8-12.

Howell EM and Blondell B. International infant mortality rates: bias from reporting differences. *Am J Public Health*. 1994;84:850-852.

Hutchkins V, Kessel S, Placek P. Trends in maternal and infant health factors associated with low birth weight, US, 1972 and 1980. *Pub Health Reps*. 1984:99:162-172.

International Bank for Reconstruction and Development/The World Bank. *Population Change and Economic Development.* New York: Oxford University Press, 1984.

Jepson HA, Talashek ML, Tichy AM. The Apgar Score: evolution, limitations, and scoring guidelines. *Birth.* 1991;18:83-92.

Jones E and Forrest J. Underreporting of abortion surveys of US women: 1976-1988. *Demography.* 1992;29:113-126.

Joyce TJ. Self-selection, prenatal care, and birth among blacks, whites and Hispanics in New York City. Paper presented at the 7th annual meeting of the Association of Health Services Research and the Foundation for Health Services, Research, June 17-19, Arlington, Virginia.

Kennedy RD, Deapen RE. Differences between Oklahoma Indian infant mortality and other races. *Public Health Rep.* 1991;106:97-100.

Kessel SS, Kleinman JC, Koontz AM, Hogue CJR and Berendes HW. Racial differences in pregnancy outcomes. *Clinics in Perinatology.* 1988;14:745-754.

Kircher T, Nelson J, Burdo H. The autopsy as a measure of accuracy of the death certificate. *N Engl J Med.* 1985;313:1263-1269.

Kleinman JC. Underreporting of infant deaths: then and now. *Am J Public Health.* 1986;76:365-366.

Kleinman JC and Kessel S. Racial differences in low birthweight. *N Engl J Med.* 1987;317:749-753.

Kogan MD, Alexander GR, Kotelchuck M, and Nagey DA. Relation of the content of prenatal care to the risk of low birth weight: maternal reports of health behavior advice and initial prenatal care procedures. *JAMA.* 1994a;271:1340-1345.

Kogan MD, Kotelchuck M, Alexander GR, and Johnson WE. Racial disparities in reported prenatal care advice from health care providers. *Am J Public Health.* 1994b;84:82-88.

Kotelchuck M. The mismeasurement of prenatal care adequacy in the US and a proposed alternative two-part index. Paper presented at the annual meetings of the American Public Health Association, New Orleans, Louisiana, October 18-22, 1987.

Kotelchuck M. An evaluation of the Kessner Adequacy of Prenatal Care Index and a Proposed Adequacy of Prenatal Care Utilization Index. *Am J Public Health.* 1994a;84:1414-1420.

Kotelchuck M. The adequacy of prenatal care utilization index: its US distribution and association with low birthweight. *Am J Public Health.* 1994b;84:1486-1488.

Kramer MS, McLean FH, Boyd ME and Usher RH. The validity of gestational age estimation by menstrual dating in term, preterm, and postterm gestations. *JAMA.* 1988;260:3306-3308.

Lambert DA, Strauss, LT. Analysis of unlinked infant death certificates from the NIMS project. *Pub Health Rep.* 1987;102:200-204.

Leveno KJ, Cunningham FG, Roark ML, Nelson SD, and Williams ML. Prenatal care and the low birth weight infant. *J Obstet Gynecol.* 1985;66:599-605.

Lubchenco LO, Hansman C and Boyd E. Intrauterine growth in length and head circumference as estimated from live births at gestational ages from 26 to 42 weeks. *Pediatrics.* 1966;37:403-408.

Luke B, Keith LG. The United States standard certificate of live birth: a critical commentary. *J Reprod Med.* 1991;36:587-591.

MacFarlane A. Commentary: revised U.S. certificate of birth: a view from England. *Birth.* 1989;16:193-195.

MacGregor SN, Keith LG, Chasnoff IJ, Rosner MA, Chisum GM, Shaw P and Minogue JP. Cocaine use during pregnancy: adverse perinatal outcome. *Am J Obstet Gynecol.* 1987;157:686-690.

Malloy MH, Kleinman JC, Garland GH, Schramm WF. The association of maternal smoking with age and cause of infant death. *Am J Epidemiol.* 1988;128:46-55.

McClain PW, Sacks JJ, Froehlke RG, Ewigman BG. Estimates of fatal child abuse and neglect, United States, 1979-1988. *Pediatrics.* 1993;91:338-343.

McCormick M. The contribution of low birth weight to infant mortality and childhood morbidity. *N Engl J Med.* 1985;312:82-89.

McCormick M, Brooks-Gunn J, Workman-Daniels K, Turner J and Peckham G. The health and development status of very low birth weight children at school age. *JAMA.* 1992;267:2204-2208.

Minton SD, Seegmiller RE. An improved system for reporting congenital malformations. *JAMA.* 1986;256:2976-2979

Myers SA and Gleicher N. A successful program to lower cesarean-section rates. *N Engl J Med.* 1988;319:1511-1516.

Nagey DA. The content of prenatal care. *Obstet Gynecol.* 1989;74:516-528.

Nakamura RM, King R, Kimball EH, Oye RK, Helgerson SD. Excess infant mortality in an American Indian Population, 1940-1990. *JAMA.* 1991;266:2244-2248.

National Center for Health Statistics. Advance Report of Final Natality Statistics, 1991. Monthly Vital Statistics Report 1991;42 (Suppl) Hyattsville, Maryland: Public Health Service.

National Center for Health Statistics. *Health, United States, 1993.* Hyattsville, Maryland: Public Health Service, 1993.

National Center for Health Statistics, LA Fingerhut and JC Kleinman: Comparability of reporting between the birth certificate and the 1980 National Natality Survey. Vital and Health Statistics, Series 2, No. 99. DHHS Pub. No. (PHS) 86-1373. Public Health Service. Washington. U.S. Government Printing Office, November, 1985.

National Center for Health Statistics, Taffel SM: Maternal weight gain and the outcome of pregnancy. Vital and Health Statistics, Series 21, No. 44. DHHS Pub. No. (PHS) 86-1922. Public Health Service. Washington. U.S. Government Printing Office, June, 1986.

National Center for Health Statistics, Tolson GC, Barnes JM, Gay GA, Kowaleski JL: The 1989 revision of the U.S. Standard Certificates and Reports. Vital and Health Statistics, Series 4, No. 28. DHHS Pub. No. (PHS) 91-1465. Public Health Service. Washington. U.S. Government Printing Office, June, 1991.

Oates RK, Forrest D. Reliability of mothers' reports on birth data. *Aust Paediatr J.* 1984:20:1858-1860.

Overpeck MD, Hoffman HJ, Prager K. The lowest birth-weights and the US infant mortality rate: NCHS 1983 linked birth/infant death data. *Am J Public Health.* 1992;82:441-444.

Papile LA, Munsick-Bruno G, and Schaefer A. A relationship of cerebral intraventricular hemorrhage and early childhood neurologic handicaps. *J Pediatr.* 1983;103:273.

Parrish KM, Holt VL, Connel FA, Williams B, LoGerfo JP. Variations in the accuracy of obstetric procedures and diagnoses on birth records in Washington state, 1989. *Am J Epidemiol.* 1993;138:119-127.

Partin MR. Explaining race differences in infant mortality in the United States. Doctoral dissertation, University of Wisconsin-Madison, 1993.

Partin MR and Palloni A. Accounting for recent increases in low birth weight among African Americans. Paper presented at the annual meetings of the Population Association of American in Miami, Florida, May 1994.

Patrick et al. The validity of self-reported smoking: a review and meta-analysis. *Am J Public Health.* 1994;84:1086-1093.

Perrin EB. The cooperative health statistics system. *Health Servicies Rep.* 1974;89:13-15.

Petitti DB, Hiatt RA, Chin V, and Crougan-Minihane M. An outcome evaluation of the content and quality of prenatal care. *Birth.* 1991;18:21-25.

Piper JM, Mitchel EF, Snowden M, Hall C, Adams M, Taylor P. Validation of 1989 Tennessee birth certificates using maternal and newborn hospital records. *Am J Epidemiol.* 1993;137:758-68.

Placek P. Maternal and infant health factors associated with low infant birth weight: findings from the 1972 National Natality Survey. In DW Reed and FJ Stanley, eds., *The Epidemiology of Prematurity.* Baltimore: Urban and Schwartzenburg, 1977.

Pow[illegible]-Griner E and Rogers RG. Strong compounding effects of multiple risks during pregnancy on birth weight and infant survival. *Soc Sci Res.* 1987:;72:49-54.

Pritchard JA and Macdonald PC. *Williams Obstetrics.* 16th ed. New York: Appleton-Century-Crofts, 1980.

Puffer PR and Serano CV. Patterns of mortality in childhood: report of the intra-American investigation of mortality in childhood. Washington, DC: Pan American Health Organization Scientific Publication, 1973.

Querec LJ. Comparability of reporting on birth certificates and national natality survey questionnaires. Vital and Health Statistics: Series 2. No. 83. DHEW publication No. (PHS) 80-1357. Public Health Service. Washington DC, U.S. Government Printing Office, 1980.

Ramey CT, Stedman DJ, Borders-Patterson A, and Mengel W. Predicting school failure from information available at birth. *Am J Mental Deficiencies.* 1978;82:525.

Rogers R. Ethnic and birth weight differences in cause-specific infant mortality. *Demography.* 1989;26:335-344.

Romm FJ and Putnam SM. The validity of the medical record. *Medical Care.* 1981;19:310-315.

Rosenberg HM. Improving Cause of Death Statistics. *Am J Public Health* 1989;79:563-564.

Rosenzwieg MR and Schultz TP. Estimating a household production function: heterogeneity, the demand for health inputs, and their effects on birth weight. *J Politic Econ.* 1983;91:723-746.

Rosett HL and Weiner L. *Alcohol and the Fetus: A Clinical Perspective.* New York: Oxford University Press, 1984.

Salmi LR, Dabis F, McKinley T. Quality of death certificates: studying or burying. *Am J Public Health.* 1990;80:751-752.

Sanderson M, Placek PJ, Keppel KG. The 1988 National Maternal and Infant Health Survey: Design, content and data availability. *Birth.* 1991;18:26-32.

Saugstad LF. Weight of all births and infant mortality. *J Epidemiol Community Health.* 1981;35:185.

Schieber G, Pullier J, Greenwald LM. Health care systems in twenty-four countries. *Health Affairs.* 1991;10:22-38.

Schottenfeld D, Eaton M, Sommers SC, Alonso DR, Wilkinson C. The autopsy as a measure of accuracy of the death certificate. *Bull NY Acad Med.* 1982;58:778-794.

Shah FK and Abbey H. Effects of some factors on neonatal and postneonatal mortality. Analysis by a binary variable multiple regression method. *Milbank Mem Fund Quarterly.* 1971;49:33-57.

Shapiro S, McCormick MC, Starfield BH, Krischwe JP, and Bross D. Relevance of correlates of infant deaths for significant morbidity at 1 year of age. *Am J Obstet Gynecol.* 1980;136:363-373.

Shryock HS, Siegel JS. *The Methods and Materials of Demography.* Orlando, Florida: Academic Press, Inc., 1976.

Stewart A, Turcaro D, Rawlings S, et al. Outcome for infants at a high risk of major handicap. In: Major Mental Handicap: Methods and Costs of Prevention. CIBA Foundation Symposium 59, Amsterdam: Elsevier/North Holland, 1978.

Susser MW, Marolla FA, and Fleiss J. Birthweight, fetal age, and perinatal mortality. *Amer J of Epidemiol.* 1972;96:197-204.

Taffel SM and Keppel KG. Advice about weight gain during pregnancy and actual weight gain. *Am J Public Health.* 1986;76:1396-1399.

Taffel SM, Ventura SJ, Gay GA. Revised U.S. certificate of birth--new opportunities for research on birth outcomes. *Birth.* 1989;16:188-193.

Tilley BC, Barnes AB, Bergstrahl E, Labarthe D, Noller K et al. A comparison of pregnancy history recall and medical records: implications for retrospective studies. *Am J Epidemiol.* 1985; 121:269-281.

Tompkins ME et al. The risk of low birthweight. Alternative models of neonatal mortality. *Amer J of Epidemiol.* 1985;122:1067-79.

U.S. Public Health Service. Caring for Our Future: The Content of Prenatal Care. Washington, DC: Dept of Health and Human Services; 1989.

U.S. National Office of Vital Statistics: Vital Statistics of the United States, 1975. Washington, DC: US Government Printing Office, 1978, Volume 1:1-36.

Westwood M, Kramer MS, Munz D, Lovett JM, and Watters GV. Growth and development of full-term nonasphyxiated small-for-gestational-age newborns: follow-up through adolescence. *Pediatrics.* 1983;71:376.

Wilcox AJ and Horney LF. Accuracy of spontaneous abortion recall. *Am J Epidemiol.* 1984;120:727-33.

Wilcox A and Russell I. Why small black infants have a lower mortality rate than small white infants: the case for population-specific standards for birth weight. *J Pediatr.* 1990;116:7-10.

Williams RL, Binkin NJ, Clingman EJ. Pregnancy outcomes among Spanish surname women in California. *Am J Public Health.* 1986;76:387-391.

Wise PH. What you measure is what you get: prenatal care and women's health. *Am J Public Health.* 1994;84:1374-1375.

Wright L. One drop of blood. *The New Yorker.* 1994; July 25:46-55.

Zahniser C, Halpin G, Hollinshead W, Kessel S, Koontz A. Using linked birth and infant death files for program planning and evaluation: NIMS workshop lessons. *Pub Health Rep.* 1987;102:211-216.

Zuckerman B., et al. Effects of maternal marijuana and cocaine use on fetal growth. *N Engl J Med.* 1989;320:762-768.

HEALTH INDICATORS FOR PRESCHOOL CHILDREN
(AGES 1-4)

Barbara L. Wolfe
Department of Economics
Department of Preventive Medicine
University of Wisconsin-Madison

James Sears
Department of Economics
University of Wisconsin-Madison

Prepared for the Conference on
"Indicators of Children's Well-Being," Rockville, Maryland,
November 17-18, 1994

Draft October 24, 1994

HEALTH INDICATORS FOR PRESCHOOL CHILDREN (AGES 1-4)

Barbara Wolfe and James Sears

Per capita health care expenditures on young children are lower than those on any other age group.[1] Although preventive care is critical for preschoolers, children between the ages of 1 and 4 experience low rates of acute and chronic illnesses, and they are studied less than are their younger (infant) and older (school-aged) counterparts. Immunization rates are the one aspect of the preschoolers' health which has recently received substantial attention, an exception which is partially attributable to the sudden increase in the incidence of measles in 1990 (see, for example, Lewit and Mullahy, 1994; Goldstein, Kviz, and Daum, 1993).

Children aged 1-4 are a particularly vulnerable group. They rely almost entirely upon others (adults) to meet their needs and make decisions on their part. Over time, successive cohorts of pre-school-age children have experienced particular social and economic events that have significant implications for their development. Several changes in particular over the last two decades suggest that recent cohorts of preschoolers are not doing well: the poverty rate for children has been increasing since the early 1970s, and the proportion of children growing up in single-parent families has increased. This makes the paucity of indicators for pre-school-age children surprising.

Public opinion suggests that children's well-being is a primary public concern. According to a 1993 article by Susan Nall Bales, recent surveys tell us that

> "The public wants children to be a top priority for government spending. ...24% chose it as their top priority; ... for 61%.... it was among their top three priorities for tax dollars." ....
>
> "Children's access to health care is more important to the public than other key children's issues." ....
>
> "There is a clear mandate for government to do more for children."
>
> "Americans are so concerned about children that they will even support new taxes." (Bales, 1993, pp. 186–187.)

A major change across cohorts of preschoolers is the declining number of those with a parent at home full-time. From 1940 to 1989 the proportion of five-year-olds with a parent who could supply full-time care dropped from 84 to 52 percent (Hernandez, 1994, p. 150). Among children aged 0-5 the proportion living with a full-time homemaker decreased from 78 percent to 32 percent in four decades, from the 1940s to the 1980s (Hernandez, Table 5.2). Sixty-five percent of children under five were in day care outside of their home in 1991.

The poverty rate of children under six increased to 25.7 percent in 1992, up from 18 percent in 1966 and from a low of 16 percent in 1973. Nearly three-quarters of these children in poverty in 1992 were covered by publicly provided medical insurance (Medicaid), but only 43 percent of children whose family income was 100 to 133 percent of the poverty line were covered, and overall, 29 percent of all children under six were covered by public health insurance.

The big question that retains great importance is, What has been the impact of these changes in terms of children's well-being? The answer requires data on outcomes, which are generally more difficult to measure than inputs. If a clear, strong relationship is established between an input and an outcome, data on the input can serve as a proxy for an outcome. Unfortunately, in most cases we have not collected adequate data to establish these links in such a way as to give researchers confidence in the input-outcome relationship. Collection of such data is therefore part of the task facing those of us interested in monitoring children's well-being (Hernandez, 1994).

---

[1] While published data are tabulated for varying age groups, Evans and Friedland (1994) estimate that children between the ages of 1 and 10 have the lowest health expenditures.

In this paper we report on measures of health collected for children aged 1-4, recommend construction of three measures concerning health status and two concerning access to medical care, and argue that these dominate other possible indicators. We then discuss the steps required to obtain information for these indicators, providing some alternatives that vary with health care policy.

In the discussion that follows we make use of three criteria to judge measures of child health:

1. Variability: the ability to detect changes over time and differences across populations.
2. Validity: actual measurement of what is intended to be measured.
3. Reliability: freedom from error (and related to this, sensitivity--that is, the probability of detecting true cases).

To illustrate: the sex-adjusted mortality rate for children 1-4 is an objective and readily available statistic which is quite reliable (free of error). However, because that mortality rate is very low, it has limited variability and hence is unlikely to detect most of the changes in the health of children. Given this, it is not a very useful or valid indicator of overall child health. We also regard the cost of gathering information on the indicator as an important consideration in the task of evaluating and recommending indicators, and we place emphasis on indicators that focus on children in lower-income households.

Current Collection of Indicator Information

Four surveys collect data on the health of pre-school-age children with some degree of regularity: the National Ambulatory Medical Care Survey (NAMCS), the National Health and Nutrition Examination Survey (NHANES), the National Health Interview Survey (NHIS), and the National Medical Expenditure Survey (NMES). Only NAMCS and NHIS are collected annually; only NHIS is a long series that provides a view of changes in health over several decades. The Rand Health Insurance Study (Rand HIS) also contributes to our knowledge of children's health status, but it is not current. The National Longitudinal Survey of Youth (NLS-Y) collected data on the children of respondents to the National Longitudinal Survey, but these are children born to women of a narrow age range and hence may not be generally representative of pre-school-age children generally. All of these data sources except NHANES are household based; NHANES has data from physical examinations. Some data on NMES are also corroborated with provider-based information.

For purposes of discussion, we divide indicators of child health into the following broad categories: overall health status, medical care utilization, impairments, and other medical conditions. We also consider how some environmental factors influence child well-being through child care experiences and the incidence of accidental injuries. Overall health status may be gauged by a general measure of health, by activity limitations, days in bed, and anthropometric measures. Medical care utilization may include measures of use and measures of access or coverage. Impairments comprise physical impairments and emotional or behavioral impairments. The attached set of tables provides details on the information collected in these categories, and their sources. Only the most recent surveys are included. In addition to the six data sources mentioned above, the Pediatric Nutrition Surveillance System (administrative data), the National Hospital Discharge Survey (administrative data), the U.S. Immunization Survey (administrative data), and the Survey of Program Participation (SIPP) are referenced.

1. Overall Health Status. These measures of general health include social measures that deserve serious consideration as indicators for pre-school-age children.

a. Respondent's (parent or caretaker) impression of the overall health of the child—excellent, good, fair, or poor. This is a commonly used standard of health for all age groups. It is easy to collect and has been validated for certain older age groups (see Maddox and Douglas, 1973; Fylkesnes and Forde, 1991). It has not been well validated for this young age group. It is subjective on the part of parents/caretakers.

b. Whether the child is able to take part in play activities (or the converse, health keeps child from taking part in ordinary play). These questions focus on whether the child can participate in normal activities for his or her age. Play is likely to be the activity that would differentiate children of this age, but physical location (e.g., urban vs. rural, small apartment, or dangerous neighborhood) may influence opportunities and therefore affect responses.

c. Anthropometric measures include the child's height, weight, and weight for height. These are objective measures and hence are attractive. They are normally collected as part of a physical examination and are included in patient records. The National Center for Health Statistics has established standards for height by age and weight for height that can be used to capture such significant deviations as "stunted" or "wasted," meaning that a child's height for age or weight for height is less than the fifth or tenth percentile. These indicators are typically used for comparisons across races (see, for example, U.S. Department of Health and Human Services, 1986, p. 22). The chief limitation is that they are not sensitive to most changes in health status. Nevertheless, they do provide some indication of the well-being of poor children relative to other children. Using NHANES, for example, researchers find that among two- to five-year-old boys (girls), those in poor families are about twice (thrice) as likely to be stunted as those in nonpoor families (Montgomery and Carter-Pokras, 1993).

Most of the other measures described on the table are poorer indicators of child health. Those related to perceived vulnerability and resistance to illness have little appeal as measures of overall health because they depend on the parents' expectations (norms), the child's exposure to illness, and the presence of siblings. Bed days or preschool days missed depend not only on the child's health but also on the parents' normal activities (i.e., their opportunity cost of keeping the child in bed or at home). Bed days or days at home may reflect the proportion of mothers who work rather than the child's health. A parent's occupation may also influence reported bed or school days missed. This class of indicators is neither valid nor reliable.

2. Medical Care Utilization. Use of medical care is among the most commonly collected data on health. In this category we include ambulatory or outpatient visits, hospitalization, and insurance coverage. Medical care is an input into the production of health; at best, it serves as a health proxy. More use may be associated with poorer health (greater need for care), yet it may also indicate adequate access, leading to improved health. For a utilization measure to act as a valid health indicator, one of these effects must clearly dominate the other.

Two measures of ambulatory care visits are often encountered: whether over some specified period the child has seen (or had any form of contact with) a provider and whether the child has a regular source of care. Unless the role of "well child" visits is understood, these measures alone cannot be viewed as valid indicators of good or poor health. When data on diagnoses or specific health conditions are collected, the utilization of providers for particular conditions may provide a more sensitive and valid set of health proxies. Choosing an appropriate recall period poses a dilemma for all such data: poor recall of long-ago events may lead to unreliable measures, but a short reference period limits variability. Diagnostic-specific information is likely to convey useful information regarding access to medical care; however, the small proportion of children with any particular diagnosis limits its role as an overall indicator of child health. (That is, it has limited validity as a measure of overall child health.)

Data on hospitalizations include number of inpatient stays, length of stay, and rate of hospitalization by diagnosis. These measures tend to be reliable and easy-to-collect indicators of poor health, but they contain limited information. Too few children experience hospitalizations in any year for these data to provide a comprehensive measure of child health. However, a comparison of hospital utilization to physician utilization may be informative. Children in poor families have been observed to use less ambulatory care and more hospital care than children in wealthy families, suggesting that they do not receive care until later in the course of their illness (U.S. Department of Health and Human Services, 1989).

Insurance coverage is another possible measure of child health. It is commonly collected, has varied substantially over time, and has historically revealed striking differences among racial and income groups. Studies have repeatedly shown that insurance coverage is linked to utilization, so it does capture an important factor that is likely to influence access to care (Manning et al., 1987).

Vaccinations prevent illness; they are easier to observe than the incidence of illness. Information on them is collected both from individuals answering questionnaires and from provider surveys. The goal of immunization is the prevention of illness, and the success rate is extremely high. As such it is a potentially useful indicator of child health and one that has historically exhibited dramatic variation. If based on parental responses, reliability is limited, however, and validity over time depends on an unvarying recommended immunization schedule.

3. Impairments. A set of health measures that is likely to have long-run impacts on school performance and acceptance by peers is that concerning impairments, including significant problems with regard to sight, hearing, development, and significant physical impairment. Development includes measures of delay in growth, the presence of a learning disability, mental retardation, and whether a provider has diagnosed the child as having an emotional, developmental, or behavioral problem. This set of measures may not be reliable; that is, it may have a built-in bias. If a child does not go to a medical care provider or a psychological testing group, the parents may be ignorant of their child's condition and hence are unable to convey the "true" information on development. The remaining impairment indicators are likely to be reliable and valid, but the low incidence of impairments limits their ability to convey significant changes in the health status of the 1- to 4-year age group.

4. Other Medical Conditions. Another set of indicators concerns acute and chronic conditions. Chronic measures that are likely to have long-term and significant consequences for a child include heart problems, diabetes, frequent diarrhea or colitis, and significant allergies. A problem with these measures is that they are not adjusted for the severity of conditions. Another problem is that parents are more likely to report chronic conditions when they have been diagnosed, and diagnosis requires contact with a provider. One could imagine that reported health status might appear to deteriorate (the number of conditions might grow), when what actually occurred was a rise in physician contact and a corresponding increase in the probability of diagnosis and treatment. A final drawback is that a count of conditions may be deceiving, since not all conditions are similar in their implications for child health.

Information on acute conditions is more typically included among the utilization indicators than in direct measures of acute problems. One exception is NHANES III, which will provide an indicator of iron deficiency anemia. While this measure may be useful for international comparisons, rates of anemia are expected to exhibit little variation for preschoolers within the United States[2].

5. Environmental Factors. Health status is linked to countless environmental factors, ranging from violence within the community to the quality of adult supervision. Because many of these environmental influences are viewed as "norms" by the people who experience them, they are unlikely to be fully reflected in parental assessments of child health or play limitations, nor can we hope to address every one of them individually. Instead, we include child care in our final set of indicators and try to capture the effects of other environmental factors through measures of safety and accidental injury.

Although child care is clearly not valid as a measure of overall health status, it may serve as a measure of parental time spent with children. It could also be viewed as a control for acute illness, since, in general, children are exposed to more disease in child care outside the home than at home. The questions asked are directed at child care quality, and quality may influence child well-being.

---

[2] This is not to suggest that anemia poses negligible risks for all age groups. Earl (1993) recommends that infants not receiving iron-fortified formula be screened for iron deficiency at age 9 months. He also suggests that children with such other risk factors as poverty or abuse be rescreened between the ages of 6 and 9.

The only aspect of child safety commonly addressed by surveys is the use of car safety seats; accidental injuries have the potential to capture the effects of a far wider array of environmental factors. The use of accidents as a health indicator presents some of the problems already mentioned for other types of medical conditions. Any count of accidents must combine conditions of varying degrees of seriousness. Even if a count is limited to injuries which required the attention of a physician, it will still reflect varying propensities to seek medical care. Nonetheless, rates of accidental injury are expected to have some validity for the comparison of health risks across populations.

Potential Indicators for Which Data Have Not Been Collected

We have seen that information on a wide range of child health indicators is regularly collected. We might wish to see some measures collected in a different manner, but almost no aspects of child health have been entirely ignored. As suggested above, safety is one health-related area that has received very little attention. Hazards to young children include bathing without supervision, open stairways, and poorly stored poisons and weapons. Parents' lack knowledge of the use of ipecac could also be considered as a sort of safety hazard. Accidental injuries, which have been the subject of survey questions, reflect many of the same health risks as safety hazards. Safety hazards would have the advantage of being much more common than resulting injuries, but survey respondents might be unaware of or unwilling to acknowledge such hazards.

Lead poisoning is a child health concern which has not been (and probably can never be) accurately assessed with national survey data. The Centers for Disease Control reduced the level of blood lead concentrations calling for intervention from 25 micrograms per deciliter in 1985 to 10 micrograms per deciliter in 1993 (National Research Council, 1993). Such low doses of lead are measurable, given strict quality control, but testing is generally limited to high-risk groups (i.e., pregnant women and young children among the poor).

Recommendations for Future Data Collection

Considering all available measures and the type of research that we might wish to do, the following five indicators seem desirable:

- Parental evaluation of overall health: excellent, good, fair, or poor;
- Whether or not the child can engage in normal play activities for age, and if not, whether play is limited by health or impairment;
- Whether the child is covered by health insurance, and if so, what type;
- Whether the child is vaccinated according to recommended standards for age; and
- Number of accidents requiring medical treatment (a visit to the hospital) and the cause of each such accident.

If we could add a sixth, it would be height for age and weight for height (or length) measured by a person who conducts the survey or who accompanies the primary "surveyor." This indicator has been gathered at long intervals by the NHANES. Given the expense involved in conducting this sort of survey, we cannot recommend collecting weights and heights more frequently. However, if health data should be gathered from providers for other reasons, we would like to see these statistics included.

Our first recommended indicator is the "excellent, good, fair, poor" measure that is so commonly collected across age groups. The advantages of this measure are that it is easily collected, commonly used, and readily understood (even if the difference between excellent and good is not clear, the difference between either of these and poor is quite apparent). This overall health indicator is likely to be most useful when collected in conjunction with work and income data, as in the NLS-Y; it would be useful to learn whether labor market behavior is influenced by the presence of children with poor health and whether health status varies with family income.

A second indicator with desirable properties is whether the child is limited in the kind or amount of play activities in which he or she can participate. Surveys regularly collect this information with the typical stipulation that the play limitation be the result of health or an impairment. Play activities could also be limited by space constraints or by a lack of time on the part of the caretaker. In such case the limitation would still seem to be an indicator of "poor health," though in a broad sense rather than a narrow medical one. Accordingly, we would recommend that child health surveys include one question regarding ability to engage in normal play activities for age and use a follow-up question to ascertain the role of health or impairments in causing any limitations.

Play limitations and parental categorization of health status share two major drawbacks as overall measures of health: they are dependent on the norms of the society in which they are asked, and they do not indicate whether health problems are medically treatable. The first of these disadvantages is inherent. If average health status in a society improves, expectations will rise, and overall trends in parental health evaluations over time will be invalid. However, these measures may still be valid for judging whether the health of various racial or income groups converges or diverges over time. Fylkesnes and Forde (1991) find a "striking gap" between conditions which affect self-reported health status and conditions which are medically treatable. The second drawback may be partly addressed by follow-up questions. Over the next decade, we would like to see follow-up questions developed that would inquire whether play limitations and poor or fair health were primarily the results of permanent impairments, self-limiting ailments such as colds, or treatable conditions.

Until or unless we have universal health insurance coverage, another indicator that provides useful information is whether or not a child has health insurance. Coverage is correlated with access to medical care; without coverage children may not get adequate medical attention for any health conditions they develop. A decrease in the proportion of children covered suggests deterioration in access to medical care and hence deterioration of the quality of life for young children. Parents should be asked whether the child is covered under Medicaid, a parent's coverage through employment, other private coverage, CHAMPUS or other government programs, or is not covered by any form of health insurance. It would also be useful to know whether the child has had the same coverage for the last six months. This will provide information on the extent, nature, and stability of coverage. This information is quite easy to collect and can be collected whenever parental information on coverage is asked, as in the Current Population Survey. The type of coverage conveys information augmenting the simple "yes, no" response, since access may be reduced, especially access to specialists, if coverage is provided publicly rather than privately.

A good deal of attention has been paid to the proportion of children who are vaccinated. (It is really the only measure for this age group that has received attention!) The typical vaccination issue facing researchers is what fraction of children are up-to-date on their immunizations at age two. Virtually all children eventually receive immunizations because of school entry requirements. Vaccinations represent prevention as well as access to medical care. They are our best indicator of preventive care because "well child" visits are difficult to count or clarify. A problem is that, because parental response is not reliable,[3] this measure really depends on administrative data. If a system were developed in which vaccination data were routinely reported to the Public Health Service, this information could be readily made available and serve as a form of administrative data on child health. This could be provided on a state or county basis and could be used to identify areas of significant underservice to the pre-school-age population. Although such a system could theoretically be implemented on a nationwide basis within the next decade, we believe a more realistic goal would be to make the administrative immunization data available for a large sample of counties within ten years.

Finally, we believe that accidental injuries requiring medical treatment may reflect environmental influences on health which are not captured in the other measures we have suggested. The easiest method to gather information on accidents would be by survey. Serious accidents are relatively rare, but we do not anticipate that parents would have difficulty remembering them. Accordingly, we recommend a recall period of a full year. Accident data could also be gathered from emergency room records, but one would need a reliable estimate of the number of children in the area served by the hospital in order to interpret them. Studies in New York City (Szapiro, 1989) provide evidence that children in poorer areas have more hospital admissions for poisonings, fractures, traumatic stupra, and coma. We have no such data nationwide or by age or race. A comparison of estimates could provide insights into the accuracy of parent-provided data and the feasibility of using hospital records.

---

[3] Goldstein, Kviz, and Daum (1993) found that one-third of parents who reported their children fully immunized without consulting immunization cards were incorrect. Immunization cards improved response accuracy for those who had them, but possession of a card was positively correlated with having immunizations.

Conclusion

What would we gain by collecting, on a regular basis, these indicators of child health for those aged 1-4? Parents' evaluation of their children's overall health, the proportion of children who can engage in play, and the proportion of children who had an accident that required medical treatment are all outcome measures. When disaggregated by income, race, or geographic location they will provide us with real measures of children's well-being. They will allow us to track whether children's well-being is improving and if not, among which groups. But these indicators are also useful to study the relationship between inputs and outcome. On an aggregate (county) basis we can study the relationship between each of these outcome measures and availability of medical care, of health insurance coverage, of average income, and such neighborhood factors as proportion of high school dropouts, proportion of female-headed households, and so forth. Similar studies can be done using individual data in a structural model to ask questions regarding the impact of poverty, of insurance, of parental time, parental education, and age of mother on children's health. With such data, we can learn far more regarding the determinants of child health. With such knowledge, public policy can be better directed.

The remaining two indicators of child health concern inputs (whether the child has health insurance coverage and, if so, of what type) and utilization; or vaccinations. They are indicators of access to medical care.

Public policy can influence these inputs. If we can establish a link to outcomes, we will have an important policy tool to improve the well-being of children aged 1-4.

REFERENCES

Bales, Susan N. 1993. "Public Opinion and Health Care Reform for Children." The Future of Children 3(2) Summer/Fall: .

Earl, Robert. 1993. Iron Deficiency Anemia: Recommended Guidelines for the Prevention, Detection, and Management among U.S. Children and Women of Childbearing Age. Washington, D.C.: National Academy Press.

Evans, Alison and Robert B. Friedland. 1994. "Financing and Delivery of Health Care for Children." Background paper for the National Academy of Social Insurance Advisory Committee on Reforming American Health Care Financing: Policy and Administrative Choices. Washington, D.C.

Fylkesnes, K. and O. H. Forde. 1991. "The Tromso Study: Predictors of Self-Evaluated Health: Has Society Adopted the Expanded Health Concept?" Social Science and Medicine 32(2): 141-6.

Goldstein, Karen P., Frederick J. Kviz, and Robert S. Daum. 1993. "Accuracy of Immunization Histories Provided by Adults Accompanying Preschool Children to a Pediatric Emergency Department." Journal of the American Medical Association 270(18) November: 2190-2194.

Hernandez, Donald J. 1994. America's Children: Resources from Family, Government, and the Economy. New York: Russell Sage Foundation.

Lewit, Eugene M. and John Mullahy. 1994. "Immunization of Young Children." The Future of Children 4(1): 236-247.

Maddox, G. and E. Douglas. 1973. "Self-Assessment of Health: A Longitudinal Study of Elderly Subjects." Journal of Health and Social Behavior 14: 87-93.

Manning, W. et al. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." American Economic Review 77: 251-277.

Miller, Jane E. and Sanders Korenman. 1994. "Poverty and Children's Nutritional Status in the United States." American Journal of Epidemiology 140(3 August): 233-243.

Montgomery, L. and O. Carter-Pokras. 1993. "Health Status by Social Class and/or Minority Status: Implications for Environmental Equity Research." Toxicology and Industrial Health 9(5): 729-773.

National Research Council. 1993. Measuring Lead Exposure in Infants, Children, and Other Sensitive Populations. Washington, D.C.: National Academy Press.

Owen, G. et al. 1974. "A Study of Nutritional Status of Preschool Children in the United States 1968-70." Pediatrics 53(Supp. Part II) 597-646.

Starfield, Barbara and Paul Newacheck. 1990. "Children's Health Status, Health Risks, and Use of Services." Children in a Changing Health System: Assessments and Proposals for Reform, ed. Mark J. Schlesinger and Leon Eisenberg. Baltimore: Johns Hopkins University Press.

Szapiro, N. 1989. "Children, Poverty and Hospital Care in New York City." In Poverty and Health in New York City, ed. M. I. Krasner. New York: United Hospital Fund of New York.

U.S. Department of Health and Human Services. 1981. Better Health for Our Children: A National Strategy. Washington, D.C.: Government Printing Office.

U.S. Department of Health and Human Services. 1986. Health USA. Washington, D.C.: Government Printing Office.

U.S. Department of Health and Human Services. 1989. Child Health USA. Washington, D.C.: Government Printing Office.

World Health Organization Working Group. 1986. "Use and Interpretation of Anthropometric Indicators of Nutritional Status." Bulletin of the World Health Organization 64: 929-941.

TABLE 1
Measures of Overall Health Status

Measure	Description	Source
Health status	Excellent, good, fair, or poor (respondent's impression)	NMES, Rand HIS, NHANES III, NHIS core, NLS-Y
	Excellent, very good, good, fair, or poor (physician's impression)	NHANES III
	Worry about child's health: not at all, little, some, great deal (past 3 months)	NMES
	Health caused pain/distress: not at all, little, some, great deal (past 3 months)	NMES
	Resists illness very well (definitely F, mostly F, mostly T, definitely T)	NMES
	(T or F)	NHIS child questionnaire (i.e., 1988, 1991), Rand HIS
	Never been seriously ill (definitely F, mostly F, mostly T, definitely T)	NMES
	(T or F)	NHIS child questionnaire, Rand HIS
	Catches things "going around" (definite F, mostly F, mostly T, definite T)	NMES
	(T or F)	NHIS child questionnaire, Rand HIS
	Less healthy than other children (T or F)	NHIS child questionnaire, Rand HIS
	Usually recovers quickly when sick (T or F)	NHIS child questionnaire
	Ever so sick you thought might die (T or F)	NHIS child questionnaire, Rand HIS
Play limitations	Able to take part at all in usual play activities (current)	NHIS core, NHANES III
	Able to take part at all in usual play activities (past 3 months)	NMES
	Has condition that limits play activities	NLS-Y
	Kind or amount of play activities limited by impairment or health problem (current)	NHIS core, NHANES III
	(past 3 months)	NMES
	Kind or amount of ordinary play limited by health	Rand HIS
	Health keeps from taking part in ordinary play	Rand HIS
	Needs more help than usual for age in eating, dressing, bathing, or using toilet because of health	Rand HIS
	Limited in any way by impairment or health problem	NMES
Bed days	Stayed in bed because of illness or injury (past 2 weeks)	NHIS core
	Days in bed > 1/2 day due to illness or injury (past 2 weeks and year)	NHIS core
	Days in bed > 1/2 day due to illness or injury (since last interview)	NMES
Anthropomorphic measures	Height and weight without shoes, age	NMES, NHANES III, The Pediatric Nutrition Surveillance System (low- income only), NLS-Y
	Low (5th percentile) height-for-age, low weight-for-height, and high (95th percentile) weight-for-height	Pediatric Nutrition Surveillance System
	Body measurements including head and upper arm circumferences and triceps skinfold	NHANES III

TABLE 2
Medical Care Utilization

Measure	Description	Source
Visits and calls	Seen or talked to medical doctor about child (past 2 weeks)	NHIS core
	Seen or talked to medical doctor or assistant about child (past year)	NHIS core
	Time since seen or talked to MD or assistant about child	NHIS core
	Is there a usual place (doctor's office, clinic, etc.) for routine care?	NHIS child questionnaire (i.e., 1988, 1991)
	Is there a usual place where child goes when sick or injured?	NHIS child questionnaire
	Is there a place child usually goes for routine care, advice, or when sick?	NHANES III
	Is there a particular medical person child sees when sick?	NHIS child questionnaire, NHANES III
	Is there a particular medical person who usually gives advice over phone?	NHIS child questionnaire
	Time since last routine care visit (<6 mo., 6 mo.-1 yr., 1-2 yrs., >2 yrs.)	NHIS child questionnaire
	Time since last saw or talked to health professional about (<1 yr., 1-2 yrs., ...)	NHANES III
	Doctor visits since previous interview	NMES
	Other medical practitioner visits since previous interview	NMES
	Nurse, physician assistant visits since previous interview	NMES
	Phone calls to medical doctor about child since previous interview	NMES
	Rate of visits by diagnosis category	NAMCS
Hospitalization	Outpatient visits since previous interview	NMES
	Emergency room visits since previous interview	NMES
	Different times overnight in hospital in past year	NHIS core
	Inpatient stays since previous interview	NMES
	Times overnight in hospital in life (except birth)	NHANES III
	Rate of hospitalization by diagnosis	National Hospital Discharge Survey
Insurance	Medicare, Medicaid, CHAMPUS/CHAMPVA, other public, any private, HMO coverage (current)	NMES
	Medicare, Medicaid, CHAMPUS/CHAMPVA, HMO, and any coverage (current)	NHIS insurance questionnaire (i.e., 1986, 1989, 1992)
	Medicare, Medicaid, CHAMPUS/CHAMPVA, employer private, other private coverage (current and since initial interview)	SIPP
	Medicaid and any insurance	NLS-Y
	Whether card available	NMES
	Medicaid coverage and use in past year	NHIS topic (1988, 1989, 1991, 1992)
	Medicare and other coverage	NHANES III

(continued)

TABLE 2, continued
Medical Care Utilization

Measure	Description	Source
Vaccinations	DPT, polio by mouth, red measles, german measles, mumps	NMES
	Tetanus including DPT	NHANES III
	Measles, mumps, rubella, DPT (# doses), and polio (# doses)	US Immunization Survey (1985 and earlier)
	DPT, measles, Hib, polio	NHIS topics (1991, 1992)
	Whether vaccination card available	US Immunization Survey, NHIS topics (1991, 1992)
	Vaccination up-to-date for age	*Statistical Abstracts*, using 1991 NHIS
	% immunized (measles, mumps, rubella, DPT, polio) out of people referring to immunization records	*Health United States*, using US immunization Survey

TABLE 3
Impairments

Measure	Description	Source
Sight	Blindness	NHIS child questionnaire (i.e., 1988, 1991), NMES
	Wears glasses or contact lenses	NHIS child questionnaire, NHANES III, NMES
	Difficulty seeing even with glasses	NHIS child questionnaire, NHANES III, NMES
	Recognizes familiar people at 2 or 3 feet with glasses	NMES
	Serious difficulty seeing	NLS-Y
	Ever had vision tested	NHANES III
Hearing	Deafness	NHIS child questionnaire, NMES
	Wears hearing aid	NHIS child questionnaire, NHANES III, NMES
	Difficulty hearing even with hearing aid	NMES
	Hears some things with hearing aid	NMES
	Serious hearing difficulty	NLS-Y
	Ever had trouble hearing, which lasted more than short period	NHANES III
	Ever had hearing tested	NHANES III
Development	Delay in growth or development (When? Seen doctor? Taken medicine?)	NHIS child questionnaire
	Learning disability (When? Seen doctor? Taken medicine?)	NHIS child questionnaire
	Emotional or behavioral problem lasting more than 3 months (When? Seen doctor? Taken medicine?)	NHIS child questionnaire
	Seen doctor/counselor for any emotional, developmental, or behavioral problem (When was last time?)	NHIS child questionnaire
	Learning disability, minimal brain dysfunction, hyperkinesis or hyperactivity, serious emotional disturbance, mental retardation	NLS-Y
	Doctor said mentally retarded	NHANES III
Physical impairments	Permanent impairment, stiffness or deformity of back, foot, leg, fingers, hand, or arm (specify)	NHIS child questionnaire
	Missing finger, hand, arm, toes, foot, leg (specify)	NHIS child questionnaire
	Crippled, orthopedic handicap	NLS-Y

TABLE 4
Other Medical Conditions

Measure	Description	Source
Number of conditions	Eleven chronic conditions including heart problem, anemia, allergies	NMES
	Nine chronic conditions identified by MD, including heart disease, epilepsy	NHANES III
	About 30 chronic listed, with space to enter other condition	NHIS child questionnaire
	Seven acute, including diarrhea, flu/virus, stomachache, earache, fever	NMES
	Anemia (acute)	NHANES III
Information about each condition	Had at least 3 months (If not, is it obviously permanent?)	NHIS child questionnaire (i.e., 1988, 1991)
	Days in bed >1/2 day due to condition (past 12 months)	NHIS child questionnaire
	Taken medication for condition (past 12 months)	NHIS child questionnaire
	Surgery for condition (past 12 months)	NHIS child questionnaire
	Pain or discomfort or upset (all time, often, once in while, never)	NHIS child questionnaire
	Bothered by condition (great deal, some, very little)	NHIS child questionnaire
	Saw doctor for condition (30 days acute, 12 months chronic)	NMES, NHANES III

TABLE 5
Health Effects of Environmental Factors

Measure	Description	Source
Accidental injuries	Accidents, injuries, and poisonings in past year	NHIS child questionnaire (1988 and 1991)
	Three most recent accidents	NLS-Y
	Causes of recorded accidents	NLS-Y
	Resulting conditions (e.g., broken bones, burns, poisoning)	NHIS child questionnaire, NLS-Y
	Place where accident occured (e.g., home, day care, street)	NHIS child questionnaire, NLS-Y
Safety	Use of child safety seats (most of time, sometimes, occasionally, never)	NHIS child questionnaire
	(Seldom or never, sometimes, nearly always, always)	NMES child questionnaire
Child Care	School, preschool, day care center, babysitter, day camp, relative (specify), other (time spent past 4 weeks)	NHIS child questionnaire, NLS-Y
	Whether child care usually in home	NHIS child questionnaire, NLS-Y
	Whether caretaker received special training	NHIS child questionnaire, NLS-Y
	How often main child care arrangement changed in past year	NHIS child questionnaire
	School, day care center or preschool, relative, at nonrelative home, other (typical secondary arrangement for work week)	SIPP
	Number of adults and children present during child care	NLS-Y

HEALTH INDICATORS FOR PREADOLESCENT SCHOOL-AGE CHILDREN

Barbara Starfield MD MPH
Professor of Health Policy
University Distinguished Service Professor
The Johns Hopkins University
School of Hygiene and Public Health

Prepared for the Conference on Indicators of Children's Well-Being.
Cloister Conference Center
November 17, 1994

The long and honorable tradition of public health and vital statistics in the United States has provided the country with a wealth of information on the health status of its population and on trends over time in these characteristics. With new imperatives for accountability of new health services systems, and with increasing evidence of inequity in the distribution of resources across the population, new types of data with new types of data systems are likely to be required.

This paper will first review the purposes for which health status measures are intended. Second, the different types of health status measures and the sources of data that can provide them are presented. Third, the major types of existing measures are discussed, with their strengths and limitations and the uses to which they are put. Fourth, major existing health indicators are presented along with the extent of their use. Finally, the paper discusses those measures that are likely to find most usefulness in the future. In this paper no attempt is made to review or suggest indicators that assess access to, use of, or performance of the health services system or parts of it. Such indicators of access, use, or quality, although important, are not considered "health status" indicators.

PURPOSES OF HEALTH STATUS MEASURES

There are four major purposes for health status measures:

- to characterize the health of communities and of the nation as a whole, to permit comparisons with other communities and with comparably industrialized countries in order to assess the adequacy of the health system in meeting major needs of the population.

- to compare the health of major subgroups of the population in order to detect systematic differences in health.

- to enable evaluations of the adequacy of specific health care interventions and the impact of interventions designed to improve health status.

- to serve as the basis for planning and targeting services in order to meet important health needs.

TYPES OF HEALTH STATUS MEASURES

Health status measures are of two major types: Health indicators and composites of health status that are expressed as profiles or as indices.

Health indicators are measures of specific aspects of health status that are assumed to represent the general state of health in the population. Death rates, low birthweight ratios, teenage pregnancy rates, reportable disease rates, and immunization rates are examples of health indicators.

Health status profiles are more comprehensive representations of health that are composed of several aspects (usually know as domains) that are aggregated to form a pictorial representation. They usually represent various aspects of physical ability or performance, mental and emotional characteristics, and social behaviors or interactions. Profiles are generally used to characterize individuals rather than populations, although they could be aggregated to populations.

Health indices are measures of health that assign a quantitative score to each of a number of components (either indicators or the domains of a profile) in order to derive a single score that enables rapid comparison of different population groups.

Health indicators are generally obtained from ongoing data collection on deaths, births, hospitalizations, and morbidity as reported in regular national health interview surveys (such as the National Health Interview Survey, the National Hospital Discharge Survey, and the CDC Risk Behaviors Survey).

Health profiles may be obtained either from health interviews that tap the important domains, from health information systems that include data on various domains, or from a combination of both. Since profiles are a relatively new concept in health status assessment, there are few examples of their use. Case-mix measures, which take information from health information systems in managed health systems, provide profiles of the burden of diagnosed morbidity in different population groups served by these organizations. The Johns Hopkins Ambulatory Care Case Mix System [Starfield, 1991], for example, has demonstrated generally similar profiles of health among individuals served by large employer-based health systems but heavier burdens of morbidity experienced by populations enrolled in Medicaid plans. Over the past decade, profiles of health of children that are comparable in concept to these developed earlier for adults (such as the Sickness Impact Profile and the SF-36) [Bergner et al 1981; Stewart & Ware 1992] are receiving attention. For example, the Child Health and Illness Profile-Adolescent Edition (CHIP-AE) [Starfield et al 1992; Starfield et al 1995] is currently being used by a variety of health plans and health services researchers in their attempts to characterize the health of populations with whom they are concerned.

Health indices are generally calculated from a set of health indicators, although they are equally amenable to use with data collected from special data collection efforts. Examples of health indices are QALYS -- Quality of Adjusted Life Years Scale [Kaplan et al 1987; Stein et al 1990].

Strengths of the health indicator approach include the widespread availability of some that are relatively easy to obtain, the generally standard way in which they are obtained, and their demonstrated usefulness in documenting systematic differences in health across different populations. For example, the relatively poor position of the United States among western industrialized nations is readily demonstrated by the use of several standard health indicators, and the disparities across the nation are greater the younger the age group that is compared [Starfield, 1993]. The United States ranks last among 11 comparable nations with regard to percent of births that are low birthweight, last in neonatal mortality, eighth in postneonatal mortality, and eleventh in infant mortality as a whole. It ranks fifth to seventh, depending on the particular age and sex group, among seven comparable nations, in child death rates resulting from accidents and injuries, and it ranks fourth to fifth, depending on age and sex group, among the same seven countries ranked for death rates resulting from medical causes. Rates for indicators in adulthood, including age adjusted death rates at age 20, years of potential life lost before age 65 (which also includes preventable deaths in infancy and childhood), and age-adjusted death rates, show generally similar poor performance (although not as large as in infancy and childhood), whereas indicators of health at age 65 place the US about midway in the rankings. It is only at age 80 that the US position approaches top ranking.

The limitations of indicators as the major method for characterizing the health of populations have to do with the policy decisions that they generate, which often are directed at the development of categorical programs to address the particular problem reflected in the indicator. As a result, US health policy is often designed to address, in piecemeal fashion, the deficiencies in care associated with the particular indicator: immunization campaigns for low immunization rates or funds for targeted prenatal care programs where low birthweight ratios are high. That is, performance on an indicator is often interpreted as a deficiency in that particular aspect of the health system rather than as a reflection of a more generalized problem that is also influencing other but unmeasured health characteristics. As a result, policy decisions often provide piecemeal solutions to a more widespread problem with the organization and financing of services. For example, low birth weight ratios are usually interpreted as an effect of poor access to prenatal services when, in reality, they may be a result of poor access to comprehensive primary health care services long antedating pregnancy.

The profile approach is designed to remedy the limitations of the indicator approach. Population groups that are found to be at a disadvantage across a range of domains can be identified and targeted for the enhancement of programs that would comprehensively address the myriad of problems that are concentrated in those populations. Moreover, profiles make it easier to detect interrelationships between different areas of health and thus to help in the elucidation of factors that predispose to poor health or, conversely, enhance the likelihood of good health. Comprehensive planning for services is facilitated and assessment of impact is more focussed on general areas rather than on specific indicators of health that may or may not be representative of health in general.

The profile approach is limited in that there are few existing instruments that have been tested for reliability and validity, although some are currently being planned or tested. Second, there is little precedent for the use of this type of measure on a widespread scale, and little understanding of its potential. Methods for assuring comparability of data collection do not yet exist and there are no well developed methods for aggregating individual profiles into community profiles.

The strength of the index approach is its conceptual simplicity. Different countries or different communities can be scored, with higher scores representing a different level of health status than lower scores. Such an approach might be particularly useful when the interest is in documenting differences in health rather than their causes. Limitations of the index approach are assumptions that intervals between successive item scores represent equivalent differences in health. One approach to overcoming this limitation is to weight component scores for their perceived importance, either by expert judgments or by consumer valuations [Patrick and Bergner 1990]. Another limitation is that a single score gives no information on specific types of deficits in health status; in order to inform policy decisions, subsequent exploration of components of the index is required for this purpose.

CURRENT HEALTH INDICATORS

Table 1 lists the four major sources of health indicators and the particular indicators that they produce.

Vital statistics have the longest history of use, and have the advantage that standard definitions are in place not only nationally but also, for the most part, internationally. This source of data provides information on death rates, by age and race, for ICD-coded causes of death, which can be aggregated to produce the categories of interest.

Data on hospital discharges, by coded cause of hospitalization, have been available from a sample of U.S. hospitals and for all hospitals in some states for several years. Data in these information systems identify health problems that should have been prevented by adequate ambulatory care.

Interview data have been collected in the United States for almost 40 years and some studies of reliability and validity were carried out during the early years of their development. When conducted under the aegis of the National Center for Health Statistics, methods of administration are standardized, with good quality control. Also, analysis generally follows a standard pattern which facilitates comparisons over time when the questions are the same (as they usually are). Computerized entry of data at the point of its collection speeds analysis time so that information from the surveys is available more quickly than in the past. Interview data yields information on reported chronic conditions, reported limitations of activity associated with these conditions, reported restriction of activity associated with acute illnesses, reported completeness of immunizations, reported health behaviors, and reported physical fitness. The Child Health Supplement also elicits some information on emotional and behavior problems. The major disadvantage of interview surveys is the unknown reliability and validity of information obtained by self-report, particularly when the survey has not been independently validated.

Examination data, as obtained by the NHANES (National Health and Nutrition Examination Survey) and its predecessor HES (Health Examination Survey), provide information on the frequency of occurrence of abnormalities that are reflected in anatomical or physiological findings. These surveys generally also include selected laboratory tests that permit estimates of the prevalence of conditions such as anemia (including iron-deficiency anemia), elevated blood lead levels, and allergies as manifested by skin tests. The major problem with physical examinations is their poor reliability, even when conducted by physicians. It has been estimated that two physician examiners agree only about 15% of the time on the presence or absence of an abnormality [Starfield, unpublished data]

Data from Clinical Information Systems

Information potentially available from clinical sources includes rates of communicable diseases, cancer incidence and prevalence rates, and rates of congenital metabolic disease (such as cystic fibrosis), as well as all diagnoses that are recorded in the process of providing services. Although most existing ambulatory care health systems have not coded diagnoses made by health care providers, it is likely that this situation will change in the future. The imperative of managed care organizations to monitor utilization and quality of care is generating interest in the development and application of case-mix measures that depend upon ICD-coded diagnoses.

Table 2 lists those indicators that are most commonly used according to the aegis under which they have been collected. The US National Health Surveys, Canadian health surveys, and two major US states are represented, as are a Canadian province, the Organization for Economic and Community Development (Europe), and the compilations prepared by the Bureau of Maternal and Child Health (MCH) and by the Annie Casey Foundation (Kids Count). Of the data collection efforts, only the vital statistics system and the US National Health Interview Survey (NHIS) are ongoing on a regular basis although the NHIS Child Health Supplement (the source of much of the indicated information) is collected sporadically. The US National Health and Nutrition Survey is irregularly periodic. (In fact, there have been only three such surveys since their inception in the early 1960s.) The compilations (MCH and Casey Foundation) depend on the availability of other sources of information.

Although other types of data are often obtained, only those in the table are population-based. Other indicators are derived from services data and therefore cannot be generalized to produce population rates. These include but are not limited to such indicators as rates of serious behavior problems in schools (whose populations do not include individuals excluded from or otherwise not in public schools as a result of behavior problems), and manifestations of under-nutrition deriving from individuals seen in facilities such as WIC clinics. Data based upon use of health-related facilities may systematically underestimate the frequency of problems in the population because they exclude individuals who are not receiving services even though they may need them; they are also unrepresentative of whole populations because they include information only on population groups eligible for their services.

The amount of information on pre-adolescent children is far less than that available for infants and preschool children; for the latter population group it is common to have information on neonatal and postneonatal mortality rates, low birth weight rates, and immunization rates, in addition to the types of information available for older children. However, a variety of types of data is at least potentially available, which makes it possible to accomplish some of the aims of health status indicators IF the data were consistently and regularly collected.

EXAMPLES OF THE USE OF HEALTH INDICATORS

a. International Comparisons

Table 3 provides some international comparisons of death rates and rates of activity restriction and limitation as published by the National Center for Health Statistics and the OECD, respectively. This information derives from special studies and no time trends are available. However, the table shows the potential of such data if they were periodically available.

b. Comparisons by Socioeconomic Status

Data by family income or parental education are consistently available only from national health interview and national health examination surveys. As a result, many of the indicators in Table 2 that are obtained by other means do not allow for comparisons by social class. Table 4 present information derived from special analyses of data in the National Health Interview survey. In contrast to the data from the survey itself, which are published by income groups, social class is categorized into three groups (poor, near-poor, and non-poor).

c. Time trends

Figure 1 shows time trends in hospitalizations by three causes, for children under age 15. (Similar graphs could be developed for particular age groups of interest.) Since hospitalizations provide utilization data, inferences to population rates of problems is fraught with potential bias. However, if access to inpatient care is generally available to all members of the population, and if hospital admission policies do not change over time, trends in hospitalization rates from different causes could be considered to reflect the frequency of these problems in the population. In Figure 1, rates of hospitalization for tonsillectomy and adenoidectomy fell over time, most likely as a result of changing hospital admission policies and medical practice rather than changes in the frequency of disease. On the other hand, rising rates of admission for the diagnosis of asthma probably are a reflection, at least in part, of increasing morbidity from asthma, because the data are consistent with rising rates as obtained by other methods. Figure 2 also shows time trends, in this case for limitations of activity resulting from chronic illnesses, as obtained from the National Health Interview Survey.

SUGGESTIONS FOR THE FUTURE

Table 5 presents a summary of health indicators recommended for preadolescent school-age children, according to 12 criteria. The first 11 of these criteria are derived from Moore [1994]; the twelfth takes into consideration the likelihood with which the indicator directly reflects health system characteristics that are amenable to change or the extent to which it provides at least the potential for elucidating the relationship between the cause of the health concern and its manifestation. The thirteenth addresses the potential for international comparisons. The indicators reflect a reasonably broad spectrum of health status, although neither mental health problems nor states of risk and resilience (characteristics of health that influence, negatively or positively, subsequent health) are well represented.

Four indicators are recommended given the current state of availability and feasibility:

- death rates, from vital statistics, presented in total and by cause aggregated into deaths resulting from accidents and injuries and those resulting from "medical" causes.

- Limitations of activity, from the National Health Interview Survey, total and by morbidity burden.

- Hospitalizations for conditions sensitive to primary care, obtained from hospital discharge data, total and by individual ICD-coded diagnosis

- Indicator conditions, obtained from various sources as noted below.

Each of the indicators received a high rating for most of criteria; other indicators not listed on Table 2 would receive lower ratings for most of the criteria.

Hospitalizations for conditions sensitive to good primary care and therefore preventable by such care is a relatively new indicator of health status. It is potentially available universally, since it depends only on the availability of discharge data that contain ICD-coded data. Such information, while not universally available now, will become increasingly so as imperatives for cost containment and accountability increase. Figure 3 shows the potential of such an indicator for characterizing differences in health, particularly those that are amenable to medical interventions, since the data can be aggregated according to geographic areas distinguished by their social characteristics (in this case, median income). As the Figure shows, rates of admissions for all of these causes are much higher among populations living in low-income areas. The potential of this indicator to show international differences is suggested by a recent study (Casanova et al 1994), which showed that rates of admission for these types of conditions among children in Valencia Spain (where access to care is universally provided) do not vary by social characteristics of areas of residence, as they do in the United States as a whole or in specific areas that have been studied.

Death rates have the considerable advantage of being widely available for international and intranational comparisons. They have already shown their usefulness for this and other purposes; time trends are relatively easily to obtain. A major limitation of this indicator is the relative rarity of deaths in childhood, so that comparisons among population groups too small to permit stable estimates of rates are not possible. But since the data are available for each year, aggregation over a period of more than one year can make the estimates more stable and permit interpretation in populations that are otherwise too small. Another limitation is the unavailability of individual data on social class, which makes it impossible to use these rates to assess systematic differences in deaths or cause of death by class except at the ecological level (where characteristics of the place of residence are assigned to deaths in that community).

Data on limitations of activity linked to acute illness requires information from health interview surveys which currently are conducted only on national samples. These national samples permit regional estimates but not state estimates. As more and more states recognize the usefulness of health interview surveys, the capacity for data collection at the state level, and perhaps even at the sub-state level, will facilitate the collection of such information periodically. Table 6 presents information obtained from the National Health Information Survey; it combines data from the chronic conditions checklist with information about restriction of activity. The major disadvantage of this indicator is its unavailability internationally. Limitations of activity as a result of chronic conditions is also a potentially useful measure.

The recommended specific indicator conditions, while generally fulfilling all criteria to a relatively high degree, are limited by their current unavailability. Since potential feasibility of data collection varies with the indicator, each will be discussed separately.

1. Communicable diseases. While reporting systems and data compilation by the Centers for Disease Control make these indicators very useful, their potential is limited by the unavailability of associated sociodemographic characteristics and by incomplete reporting. They are particularly useful in reflecting the adequacy of the health system in providing immunizations to prevent these conditions. Therefore, efforts to improve reporting rates should be continued, and efforts should be made to obtain information about sociodemographic characteristics, either of the individual with the disease or by area of residence of the individual.

2. Iron-deficiency anemia and elevated blood lead levels. Information on both of these conditions is available from the National Health and Nutrition Examination Survey, which tests for their presence. The usefulness of such information has been demonstrated by analysis of time trends in blood levels among children in the United States over the past several decades. (Figure 4) Unfortunately there is currently no possibility of international comparisons, since other countries do not routinely collect such data. The major problem with these data is the irregularity with which the survey is conducted. It would be helpful if national policy led to greater regularity of these surveys.

3. Morbidity burdens. The imperative for accountability within new organizational arrangements for health services delivery will stimulate the development of information systems that collect information on diagnoses made during the course of clinical care. As health care organizations taken on responsibility for defined populations over period of time, there will be a need for case-mix systems that provide the basis for higher reimbursements to facilities with sicker populations. These case-mix systems are likely to be based on demonstrated morbidity as well as on age, gender, and social class. (Starfield et al 1991; Weiner et al 1991) As a result, it will be possible to characterize populations by the burdens of diagnosed morbidity. These methods characterize morbidity burdens, including those associated with mental health, as various combinations of different types of diagnoses experienced in a year. Figure 5 shows the potential of such a measure in demonstrating the general similarity of overall burdens of morbidity among children enrolled in three of the four HMOs but the higher morbidity of poor children (those enrolled in Medicaid). With the increasing sophistication of information systems, enrollment files (with sociodemographic information) and clinical data can be merged to permit the analysis of morbidity burdens by social class and other sociodemographic characteristics. This information is not likely to be available internationally, or even nationally (at least for a long time). However, efforts to begin such an approach should be encouraged and supported as investments in future health indicators for children.

Table 5a provides three additional types of indicators that are recommended for consideration, along with ratings against the criteria. The first concerns mental health problems. Since these problems are among the most common health concerns in the population, they should be included in any set of health indicators. The Child Health Supplement of the National Health Interview survey contains a set of questions directed at eliciting the frequency of behavioral and affective problems in the population of children. While research on the usefulness of these indicators either for planning or evaluation of the impact of health services is needed, their inclusion in the core set of indicators provides a more appropriate balance to the current sole focus on physical manifestation of health. The second, behaviors that influence subsequent health, is potentially available from interviews. The two particular ones provided as examples (unlocked loaded guns and television viewing) have both been demonstrated to influence health; both have been tested and found to have adequate reliability and validity [Starfield et al 1995]. The third additional indicator concern self perceptions of health, which have also been shown to be useful. Perceived well being reported as excellent, very good, good, fair or poor is a standard question in the National Health Interview Survey. Responses to this question have been shown to be predictive of subsequent health in adults, although no studies concerning children have been conducted. Responses to the question have been found to be related to social class, in children as well as adults, with more disadvantaged individuals reporting poorer health. Both self-perceived health and self-esteem have been shown to have moderate correlations with other aspects of health in the adolescent health profile [Starfield et al 1995] although studies in younger children have not yet been done.

Technical considerations

All of the suggested indicators should be produced by individual year of age, aggregated for ages 5-7, 8-10, 11-14, 15-17, and for 5-10 and 11-17, to provide information about specific developmental stages of childhood.

Presentation of information by social class would be facilitated if there were standard classifications that were adopted. Instead of income categories, or specification by poor, near poor, and non-poor, data might be aggregated according to those in the lowest 10th percentile of income in the population, those from 11-24th percentile, 25-49th percentile, and 50th or above. This would have the advantage of standardizing comparisons across population groups that differ in income because of geographic factors. Since information on the distribution of wealth in various countries is often depicted in this way, collection of data in this manner would permit analysis of data on equity in distribution of health in addition to that of social welfare.

Periodicity of information is less important than regularity of scheduling for its collection. In general, every five years (except for those items that are currently collected more frequently) seems appropriate, although new types of information systems (such as those derived from clinical facilities) should have information on-line and be very easy to produce continuously. Health examination surveys should be carried out regularly, at least once every 5-10 years.

These suggested indicators represent a reasonable and practical set for the near future. Developmental efforts recently completed or currently underway will provide, within 5-10 years, more comprehensive profiles of health to complement the indicators suggested above. Combined with other indicators that reflect the state of access to health services and their actual and perceived quality, they should move the country forward to a new generation of data systems that are better suited to planning and evaluation of societal policies and programs.

References

Annie E. Casey Foundation. Kids Count Data Book: State Profiles of Child Well-Being. Greenwich, CT. 1994.

Bergner M, Bobbitt R, Carter W, Gibson B. The Sickness Impact Profile: development and final revision of a health status measure. Med Care 1981; 19:787-805.

Canadian Institute of Child Health (CICH). The Health of Canada's Children: A CICH Profile. Ottawa, Canada 1989.

Casanova C, Starfield B. Hospitalizations of Children and Access to Primary Care: A Cross-National Comparison. International J Health Services 1995 (in press).

Center for Health Economics Research (CHER). Access to Health Care: Key Indicators for Policy. Robert Wood Johnson Foundation, Princeton NJ. November 1993.

Institute of Public Policy Management (IPPM). The State of Washington's Children. University of Washington, December 1992.

Kaplan R & Anderson J. The quality of well-being scale: Rationale for a single quality of life index. In: Quality of Life: Assessment and application. ed S Walker & R Rosser. Proc Centre for Medicine Research Workshop held at the CIBA Foundation, London 1987.

Manitoba Centre for Health Policy and Evaluation (MCHPE). Population Health: Health Status Indicators. Volume I: Key Findings. University of Manitoba, January 1994.

Maternal and Child Health Bureau (MCHB). Child Health USA '93. USDHHS Pub No. HRSA-MCH-94-1. March 1994.

Moore K. Criteria for Indicator of Child Well Being. Prepared for the Conference on Indicators of Children's Well-Being. November 1994.

Morbidity and Mortality Weekly Report (MMWR). Blood Lead Levels -- United States 1988-1991. MMWR 1994; 43(30):545-548.

National Center for Health Statistics. L. Fingerhut, 1989. Trends and Current Status in Childhood Mortality, United States 1980-85. Vital and Health Statistics. Series 3 No 26 DHHS Pub No (PHS) 89-1410, Public Health Service. Washington. U.S. Government Printing Office.

National Health and Nutrition Examination Survey (NHANES). MMWR 1994; 43(30):546.

New York State - Bureau of Child and Adolescent Health. Maternal, Child and Adolescent Profile. New York State Department of Health. Albany NY 1989.

OECD (Organization for Economic Cooperative and Development) Living Conditions in OECD Countries. Social Policy Studies No.3, Paris 1986.

Patrick D and Bergner M. Measurement of Health Status in the 1990s. Ann Rev of Public Health 1990; 11:165-83.

Starfield B, Riley AW, Green BF, Ensminger ME, Ryan SA, Kelleher K, Kim-Harris S, Johnston D, Vogel K. The Adolescent CHIP: A Population-Based Measure of Health. Med Care. 1995 (in press).

Starfield, unpublished data 1994

Starfield B, Bergner M, Ensminger M, Riley A, Ryan S, McGauhey P, Skinner A, Kim S. Adolescent Health Status Measurement: Development of CHIP. Pediatr 1993; 91(2):430-435.

Starfield B. Childhood Morbidity: Comparisons, Clusters, and Trends. Pediatrics 1991; 88(3):519-526

Starfield, B. Primary Care. J Ambulatory Care Manage 1993; 16(4):27-37.

Starfield B, Weiner J, Mumford L, Steinwachs D. Ambulatory Care Groups: A Categorization of Diagnoses for Research and Management. Health Services Research 1991; 26(1):53-74.

Stein REK, Jessop DJ. Functional status II(R). A measure of child health status. Med Care, 1990; 28(11):1041-1055.

Stewart A, Ware J. Measuring Functioning and Well-Being: The Medical Outcomes Study Approach. Duke University Press, Durham, NC. 1992.

Weiner J, Starfield B, Steinwachs D, Mumford L. The Development of a Population Oriented Case-Mix Measure for Application to Ambulatory Care. Medical Care 1991; 29(5):452-472.

TABLE 1

SOURCES OF INFORMATION FOR HEALTH INDICATORS

- A. Vital Statistics and Surveillance Data
  - Death rates by age group
  - Death rates for Injuries/Accidents, by type and aggregated
  - Death rates for Medical Causes, by type and aggregated
  - Deaths from Sentinel Conditions
- B. Hospitalizations, by diagnosis
- C. Interview Data
  - Reported Chronic Conditions
  - Reported Restrictions of activity, by nature of acute illness
  - Reported Limitations of Activity, by degree of interference with major or visual activities
  - Reported Behavior Problems
  - Reported Health Behavior
  - Reported Physical Fitness
  - Reported Completeness of Immunizations
  - Reported Overall health as excellent, very good, good, fair, or poor.
  - Health Profiles
- D. Examination Data
  - Physical Examination findings of manifested abnormalities
  - Laboratory examinations
    - Anemia
    - Elevated blood lead levels
    - Skin testing for allergies
- E. Data from Clinical Information Systems
  - Reportable diseases
    - Communicable Diseases
  - Case Registers
    - Cancer Registers
    - Congenital Metabolic Disease, e.g. Cystic Fibrosis
  - Diagnosed Morbidity/Disability
    - Diagnoses, individual and aggregated by type
    - Hospital discharges, by diagnosis

Table 2

PREADOLESCENT SCHOOL-AGE CHILD -- INDICATORS USED IN SELECTED PUBLICATIONS -- (U.S. AND CANADA)

	Bureau of Maternal & Child Health	Kids Count	New York State	Washington State	Canadian Institute of Child Health	Province or Manitoba**	OECD	US National Health Surveys & Vital statistics
Periodicity	Annual	Annual	?	?	Occasional	?	Occasional	Regular/Irregular
Age Aggregation								
Death Rates	✓	✓	✓	✓	✓		✓	✓
Injuries	✓		✓	✓	✓	✓	✓	✓
Medical Causes						✓	✓	✓
Sentinel Indicators								
Communicable Diseases			✓	✓				✓
Iron deficiency anemia								✓
Lead poisoning	✓		✓					✓
Restricted activity by acute illness								✓
Limitations of activity	✓		✓			✓	✓	✓
Mental Health					✓			
Child abuse*	✓		✓					
Pediatric AIDS*	✓		✓	✓				
Physical Fitness								
Dental Caries	✓			✓				
Behavior Problems				✓				✓
Hospitalizations, by diagnosis			✓			✓		
Obesity					✓			
Health Behaviors					✓***			✓****
Overall health†					✓			✓

* All child ages combined

** All data expressed as Standard Mortality Ratios (not age stratified)

*** e.g., physical activity, smoking, drinking, child restraints

**** seat belts

† Reported by parental respondent as excellent, very good, good, fair, poor

Note: Specifically absent from any of these published data sources are: under nutrition, health-protecting behaviors such as bicycle helmets, gun exposure, or smoke detectors, or specific diseases or conditions other than communicable diseases.

Source: MCHB 1994; NY-BCAH 1989; MCHPE 1994; IPPM 1992; CICH 1989; Annie E. Casey Foundation 1994.

Table 3

Child Health Indicators: International Comparisons, 1985.

		COUNTRY*								
		AU	CA	GB	FR	FRG	J	NE	SW	US
Deaths per 100,000 population										
age 5-9	Female	20	19	18	21	22	15	18	11	21
	Male	30	26	22	28	24	27	22	20	28
age 10-15	Female	16	20	19	19	17	13	19	15	21
	Male	29	31	29	30	23	20	29	17	35
Disability and activity limitation (age 0-15)										
Disability days (per person per year)		14	9	17				11		12
Bed Days (per person per year)		4	4					3		5
Activity restriction due to long-standing conditions (percent of population)		3	3	6		1		2		4

*Australia (AU), Canada (CA), Great Britain (GB), France (FR), Federal Republic of Germany (FRG), Japan (J), Netherlands (NE), Sweden (SW), United States (US).

Source: NCHS, Fingerhut 1989; OECD 1986

TABLE 4

CHILD HEALTH AND SCHOOL ATTENDANCE, 1991.

	School Days absent	Excellent/very good health ratings (% of population)
Poor	6.4	64.0%
Near Poor	5.0	74.7%
Non-poor	4.9	87.4%

Source: Original analysis of the 1991 National Health Interview Survey by the Center for Health Economics Research. Access to Health Care Indicators for Policy. Nov. 1992.

TABLE 5
RECOMMENDATIONS FOR INDICATORS, BY DESIRABLE CRITERIA, PREADOLESCENT SCHOOL-AGE CHILDREN

	Hospitalizations for ambulatory care sensitive conditions	Death rates total and cause, aggregatable type Ages 5-7 8-10 11-14 15-17	Limitations & Restrictions of Activity, total and by morbidity burden	Indicators Condition: Communicable Diseases, Iron deficiency anemia, Blood lead levels Morbidity burden
Comprehensive Coverage	Selected indicators represent preventable mortality, preventable morbidity, impact of morbidity on functional status			
Potential to track across ages	good	good	good	good
Clear & Comprehensible	good	good	good-moderate	good
Positive/Negative	negative	negative	negative	negative
Common interpretation	yes	yes	yes (maybe less for international comparisons)	yes
Consistency over time	yes	yes	yes-with standard instrument	yes
Reliability/Validity	yes	yes	yes	yes
Geographic Detail	yes	yes	poor, with existing data	variable
Cost Efficient	high-moderate	high	high-moderate (given periodic availability of data source)	moderate-low
Reflective of Social Goals	high	high	high	high
Adjustable for demographic characteristics/SES	only at ecological level (community of residence)	only at ecological level (community of residence)	high	variable
Relationship to health system characteristics a potential for linkage between cause & effect	high	moderate	moderate	high
Availability for international comparisons	no	yes	no	no

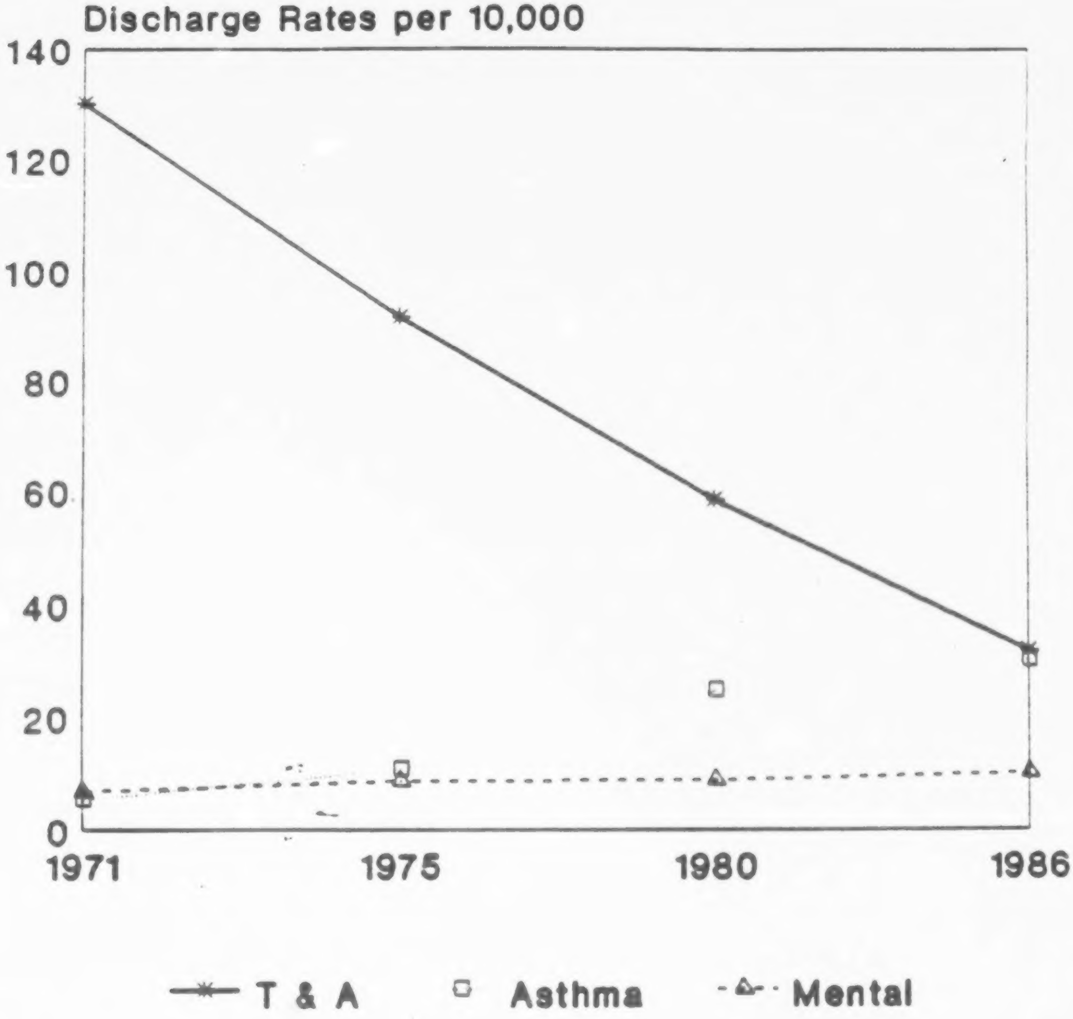
TABLE 5a

ADDITIONAL RECOMMENDATIONS FOR INDICATORS, BY DESIRABLE CRITERIA, PREADOLESCENT SCHOOL-AGE CHILDREN

	Major health risks e.g. unlocked loaded guns in household, television viewing	Sense of well being and overall health e.g. ratings of overall health and self-esteem	Behavior Problems Checklist
Comprehensive Coverage		Increases comprehensiveness	
Potential to track across ages	good	good	poor
Clear & Comprehensible	good	good	good
Positive/Negative	negative	positive	negative
Common interpretation	yes	yes	yes
Consistency overtime	yes	yes	probably
Reliability/Validity	probably	yes	yes
Geographic Detail	yes	yes	poor, with existing data
Cost Efficient	high-moderate	high-moderate	high (given periodic availability of data source)
Reflective of Social Goals	high	moderate (?)	high
Adjustable for demographic characteristics/SES	high	high	high
Relationship to health system characteristics a potential for linkage between cause & effect	moderate	moderate	moderate
Availability for international comparisons	no	no	no

Figure 1

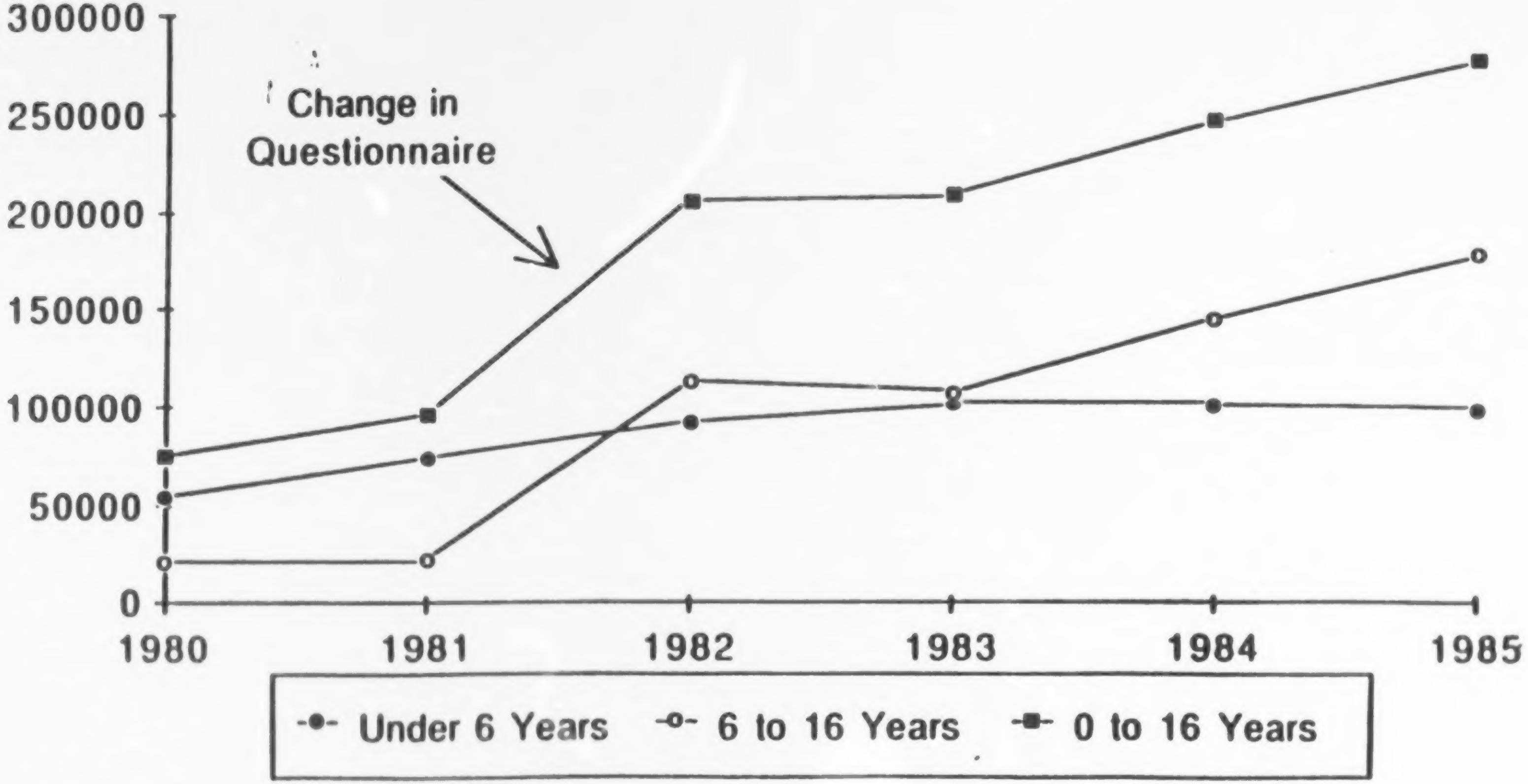
Trends in Hospitalization by Selected Causes for U.S. Children Under age 15



Source: Starfield 1991

Figure 2

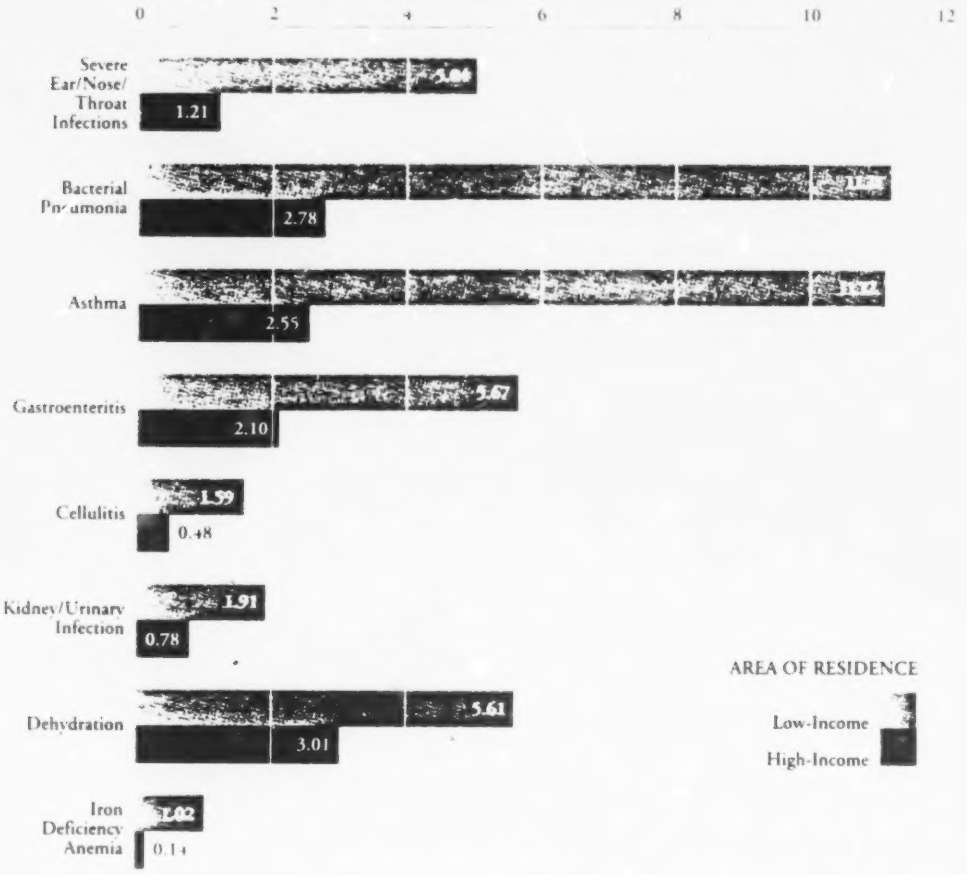
Estimated Number of Children Unable to Conduct their Major Activity due to Chronic Conditions: U.S. 1980-85



Source: Newacheck P, 1987 personal communication (as published in Starfield 1991)

Figure 3

Hospitalizations for Ambulatory Care Sensitive Conditions per 1,000 Children Under Age 5, 1989



NOTES

Data from 10 states having a total population of 95 million. High-income areas are zip codes in which fewer than 15 percent of households have incomes below $15,000; low-income areas are zip codes in which 30 percent or more of households have incomes below $15,000.

SOURCES

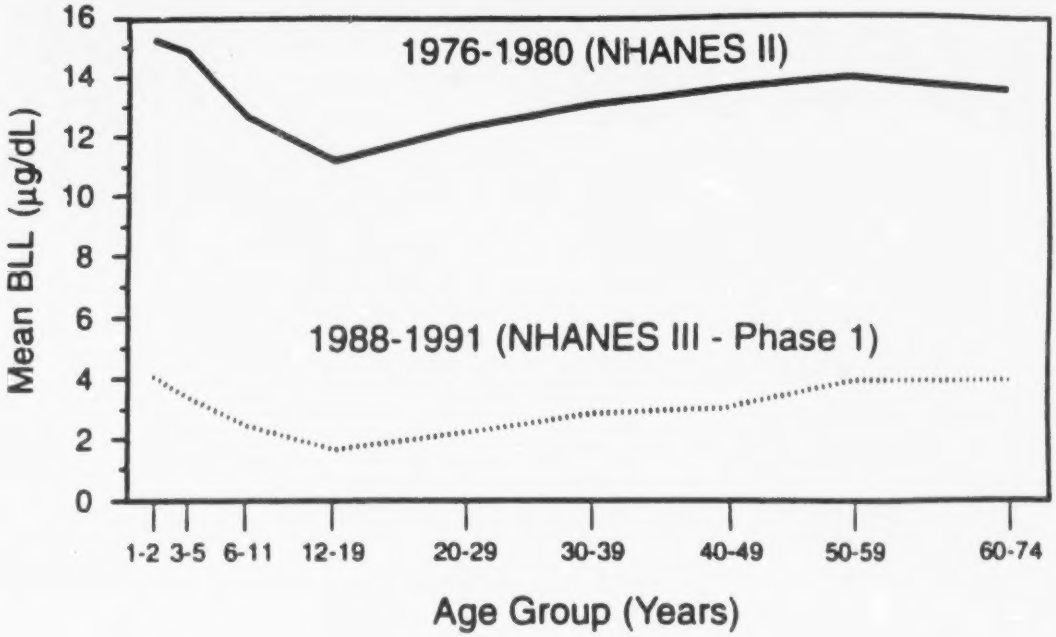
Codman Research Group. Ambulatory Care Access Project. New York: United Hospital Fund of New York.

BEST COPY AVAILABLE

Source: CHER 1993

Figure 4

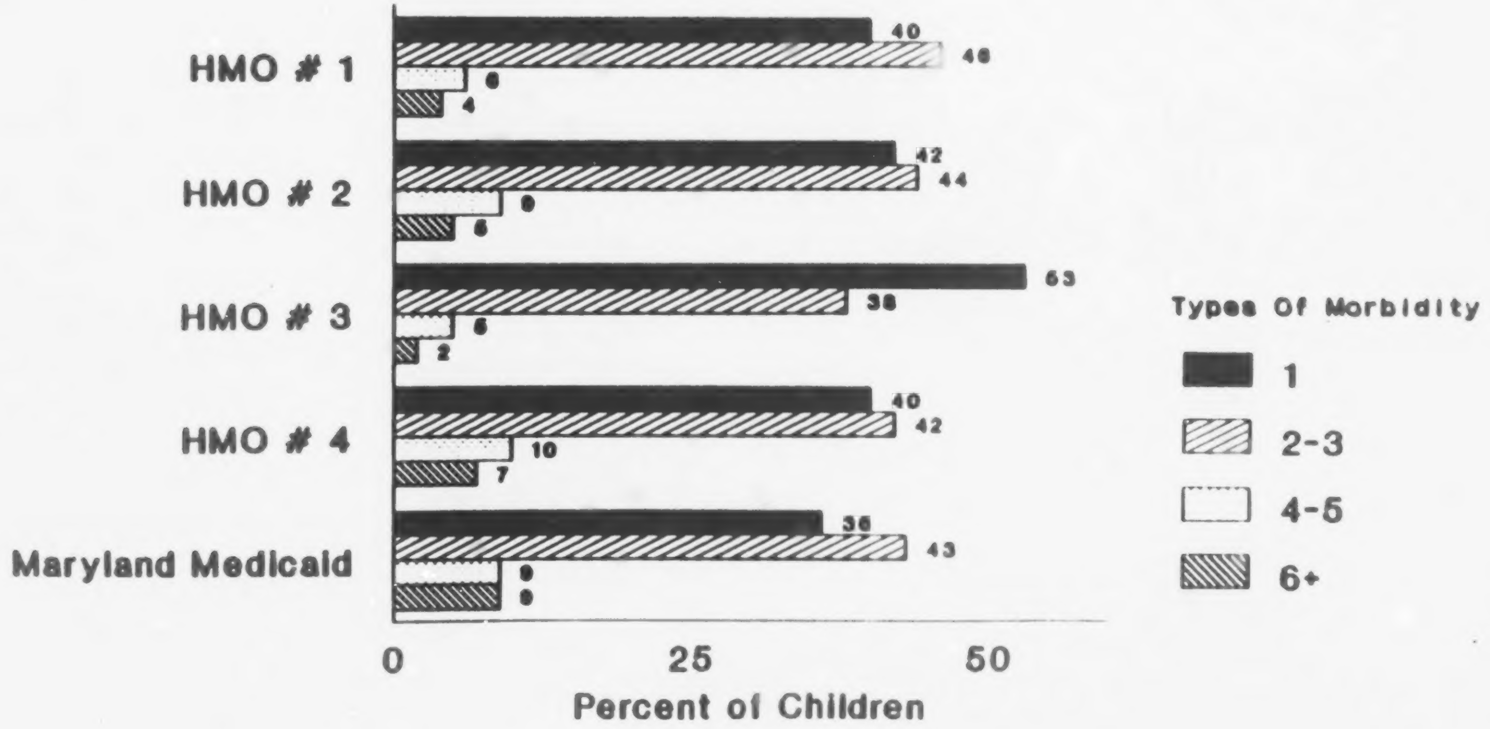
Geometric mean blood lead levels (BLLs) for persons aged <75 years, by age group — National Health and Nutrition Examination Survey (NHANES) II and III–Phase 1, United States, 1976–1980 and 1988–1991



Source: MMWR 43(30)1994

Figure 5

Percent of Children with Different Burdens of Morbidity in 5 Health Care Facilities



Source: Starfield et al 1991

Adolescent Health Indicators

Arthur B. Elster, M.D.
Director, AMA Department of Adolescent Health
515 North State Street
Chicago, IL 60610

Paper Presented at the Conference,
Indicators of Children's Well-Being

November 17-18, 1995
Rockville, MD

ADOLESCENT HEALTH INDICATORS

INTRODUCTION

A critical first step in identifying a set of indicators for assessing health and well-being is to determine the possible uses of such indicators. What are the advantages and what are the disadvantages? Above all else, we must ensure that we "do no harm".

It is reasonable to assume that health indicators measured accurately, regularly, and across a broad spectrum of the population can be a valuable mechanism for tracking progress toward achieving identified national goals. Used in this fashion health indicators can help guide program planning, research, and education.

Selected health indicators for children and adults have been used in the above manner for many years. Although there are many examples, two widely accepted indicators are those used to monitor prenatal care and pregnancy outcome, and an index used to monitor adult health risk behavior.

Cesarian section rates and percent of women who enter prenatal care in the last trimester are often used as indicators of the adequacy of prenatal care. Low birth rate, infant mortality, and whether the newborn went to an intensive care unit have been used as indicators of pregnancy outcome. These indicators meet several important criteria that have made their use widely accepted: they can be measured routinely and universally from birth certificates without additional financial cost and they have a high degree of face validity. Health advocates have used these two sets of indicators to successful lobby for increased governmental funding for obstetrical and prenatal nutrition programs. Growth of the Women, Infant, and Child (WIC) program during the period of reduction in funding for social programs that occurred in the 1980s and recent expansion of Medicaid to cover pregnancy and infant care are good examples of how health indicators can be used to promote health and well-being.

Another set of indicators has been used to monitor adult health risk behaviors. Developed by the Centers for Disease Control and Prevention (CDC), the Behavioral Risk Factor Survey includes eight behaviors linked to the ten leading causes of premature death among adults (1). State data are collected and reports are published by the CDC. This index provides a mechanism for not only tracking changes in adult preventive behavior over time, but also for comparing the health of adults among various states and regions.

Most indicators used to monitor adolescent health focus on problem behavior. Use of alcohol, drugs, and tobacco; adolescent pregnancy, live births, and abortion; and homicide comprise the majority of adolescent health indicators that are monitored and reported to the public on a regular basis. Contextual factors and health promoting behaviors are not measured as regularly as are health risk behaviors (2). Probably the two most widely used indicators of adolescent health are data from the Youth Risk Behavior Surveillance System (YRBSS) conducted by the CDC and the Monitoring the Future Survey conducted by the University of Michigan

The YRBSS was developed by the CDC in 1988 to "...identify and periodically monitor important health behaviors among youth" (3). The survey targets six behaviors: 1) behaviors that result in unintentional and intentional injuries; alcohol and other drug use; sexual behaviors that result in HIV infection, other sexually transmitted diseases, and unintended pregnancy; tobacco use; dietary behaviors; and physical activity. Surveys are conducted through most state departments of education and large local educational agencies. Representative high schools in the community are chosen and all students in these schools are surveyed. The strengths of the YRBSS are that it monitors both health risk behaviors and two health promoting behaviors; includes a national representative sample of youth; and is conducted on a relatively frequent basis. The major problem with the YRBSS is that it is conducted through state and local departments of education and, thus, is excluded from some states while other states refuse to include questions on sexuality.

Monitoring the Future is a national survey of high school seniors that has been conducted annually since 1975 by the University of Michigan's Institute for Social Research (4). Funded by the National Institute on Drug Abuse, this survey tracks alcohol and drug use attitudes and behavior among high school seniors. These findings are reported annually and have served to increase awareness of substance use among adolescents. The strengths of this survey are that it includes a national representative sample of youth and that the results have become the standard for tracking adolescent drug usage. The major drawbacks are that it only surveys students who are still enrolled in school and that it focuses on a relatively narrow range of health problems.

Surveys, such as the National Longitudinal Survey of Youth and the National Health and Nutrition Examination Survey measure a broad range of health issues including some that relate to adolescents. Since these surveys, however, are either not ongoing or else are done only periodically their use in developing adolescent health indicators may be limited. There are various other national surveys, such as the National Hospital Ambulatory Medical Care Survey and several reproductive health surveys that provide valuable information for constructing adolescent health indicators (2).

The discussion so far has been on how health indicators are used to monitor conditions selected as high national priority. Although probably unintended, health indicators can also impact society by setting standards, or at least influencing the way people think about issues. This can have both a positive and a negative influence on shaping public opinion and concern. For example, reporting on distinctions among special populations, such as racial and ethnic groups and adolescents, has had a positive effect on bringing to the public's attention the fact that our society is heterogeneous with different health care needs.

If health indicators can presumably have a positive effect on program development and on public perception, what then are the potential or real ways that indicators can be harmful to adolescents? There are at least three ways. One is the way that indicators, as described previously, can negatively influence public opinion. For example, the current use of adolescent health indicators to track problem behavior tends to distract from the many positive behaviors exhibited by adolescents. In addition, the negative and aggregate manner in which findings are reported tends to hid the fact that most adolescents are relatively uninvolved in problem behavior and that most serious problems cluster among only a sub-population of adolescents. The negative implication of indicators probably serves to further emphasize society's view that all adolescents have problems. By focusing on problem behaviors, health indicators fail to help society develop more nurturing attitudes toward youth.

A second way that use of health indicators may be problematic is data can lead to erroneous interpretations, especially in light of the atheoretical manner that indicators are often constructed. For example, for years the National Center for Health Statistics has reported children and youth data according to age categories that run counter to developmental principles. Research on adolescent pregnancy and parenthood, and on other issues, has been hampered by this approach because data hide critical age distinctions. Thus, combining data of youth 12 to 14 for purposes of reporting is logical and appropriate, while combining data of youth 15 to 19 obscures important distinctions between school-age and older adolescents.

A third way indicators are problematic is that they can not accurately reflect complex behaviors. Although select indicators may reliably measure health conditions that have discrete outcomes, such as the rate of low-birth weight infants, categorical measures are excessively reductionistic. Single health indicators can not possibly measure complex health issues that have poorly defined antecedent processes or whose meaning is abstract. This is especially problematic for adolescents in that the health of this population is reflective of factors in multiple physical, psychological, and social domains. Monitoring the rate of drinking among adolescents is a good example. Although illegal before age 21, many people in society apparently only become alarmed about drinking when adolescents are involved in motor vehicle deaths while under the influence of alcohol. By focusing predominantly on alcohol consumption, indicators as currently reported and used understate the role alcohol plays in adolescent morbidity and mortality, education and vocational underachievement, and social dysfunction.

In summary, because of rapid physical, social, psychological, and behavioral changes associated with adolescence identifying an appropriate set of indicators to measure adolescent health and well-being is a difficult task. The types of measures that could be tracked are many. Unfortunately, some of the most prominent health issues affecting adolescents have become highly politicized. In many ways, adolescents are a mirror of our society in that their behaviors mimic adult behaviors. What we dislike about adult behavior, such as infidelity, alcohol abuse, drug abuse, and excessive violence and are unwilling to take a strong stand against, we can project on our adolescents. Because of the risk that adolescent health indicators can be used punitively, great care must be taken when selecting the type of issues to measure, the ways in which the data will be analyzed, and the types of reports that will be produced.

After reviewing the ways in which they can be used, the next step in identifying a set of health indicators is to provide a working definition of health and to describe special issues of health that relate to adolescents. Assumptions will be presented that could form the foundation for identifying adolescent health indicators. Finally, a scheme for organizing health indicators will be presented along with the results of a survey of national experts regarding their choice of health indicators for adolescents.

WHAT ARE PARAMETERS OF ADOLESCENT HEALTH

Broadly defined, health is the maximal obtainable state of physical and emotional well-being. Health, therefore, is not an outcome of life, but a major resource for life. Identifying a set of indicators that measure adolescent health requires an understanding of how health is conceptualized and determined; the fact that health indicators for adolescents are both an outcome and an antecedent; and that the current nature of adolescent morbidity necessitates a greater emphasis on prevention.

Health as a State of Equilibrium

Adolescents' level of health is determined by their current state of physical equilibrium with their internal and external environment and their potential to maintain that balance. Adolescents need the reserve and resources to cope with environmental influences and to keep this balance. Physical and emotional disorders; personal behaviors, such as alcohol and drug use, unsafe sexual practices, and possession of guns; family dysfunction; and dangerous community and school environment are threats to this equilibrium. Based on this concept, it is understandable why involvement in multiple health risk behaviors is a greater threat to equilibrium than a single health risk.

Using this definition, health and factors that promote health, encompass a broad band of issues. From the medical perspective, health practitioners need to expand their teaching to think more of the role that sociological factors play in influencing health. From the sociological perspective, health researchers need to broaden their concepts to include the manner and degree in which medical conditions influence a person's ability to function in society. Working groups such as this, that bring together an eclectic collection of health scientists, economists, social and behavioral scientists, and education specialists provide a good opportunity to take a comprehensive look at health and the most reliable and valid indices to measure health.

The Dual Nature of Health Indicators

Because of rapid growth and development, adolescent health indicators serve both as an outcome measure of earlier changes, as well as a measures of how well young people are preparing themselves for a healthy adult life. For example, whether or not a young adolescent participates in sexual intercourse results, in part, from earlier psychological factors. This same behavior, however, is also an indicator of future reproductive health. In addition, some indices might have immediate implications while others affect health only many years later. Understanding the dual nature of health indices for adolescents is important for determining what measure to include in a package of indicators.

Even in its simplest form, a set of adolescent health indicators would need to focus on conditions that threaten current health equilibrium as well as those that threaten future health. In a more expanded mode, the set of indicators might include factors that precede or even predict conditions that threaten health.

Changing Focus of Health Indicators

Changes in the nature of adolescent morbidity and mortality over the past several decades have resulted in greater attention directed at health risk behaviors and the prevention of these behaviors. Whereas 30 years ago most adolescent morbidity and mortality were due to natural causes, today the leading causes of death among adolescents are related to preventable, personal behaviors--motor vehicle accidents, homicide, and suicide (5). Until recently, initiatives have addressed categorical issues, such as alcohol use, unintended pregnancy, and tobacco use. Although this approach has led to important discoveries and to the growth of special interest groups for both research and services, it also has had some unfortunate consequences. Specifically, efforts highly focused on categorical conditions have led to scholarly separatism; attention that is directed at the problem, rather than on the adolescent as a person or the family and community as an integrated unit; an atheorectical approach to the analysis of adolescent health; a sensationalism of health risk behaviors that is politically polarizing and that leads society to perceive the period of adolescence as dominated by problem behavior and family discord; and an overshadowing of disease prevention at the expense of health promotion.

The measurement of adolescent health behaviors is complicated by several important developmental issues:

1. Some degree of behavioral experimentation is normal and expected. The challenge is how to use relatively simple indices that distinguish experimental, non-problematic behavior from behavior that is destructive.

2. The significance of various health risk behaviors varies by developmental age, by culture in which the adolescent lives, and by political decisions. For example, most health professionals would agree that sexual intercourse at age 12 is problematic, while intercourse at age 16 may not be problematic depending on emotional maturity and the other factors. However, within a religiously conservative community, intercourse at age 16 probably indicates a greater willingness to deviate from community standards than does the same act in less conservative communities. For most adolescent health risk behaviors, with smoking one of a few exceptions, there is a lack of clear national priorities for the goals of prevention. Because of this, the relevance of certain behaviors, such as sexual intercourse and alcohol use, varies depending on socio-political decisions.

3. Although adolescents identify similar health concerns as do adults, the priority they ascribe to these issues differs (6). Like adults and health professionals, adolescents are concerned about the leading morbidities, such as substance abuse and the consequences of sexual behavior. Unlike adults and health professionals, however, adolescents are even more concerned about problems related to appearance (i.e., weight and acne), emotional states (i.e., anxiety and depression), interpersonal relationships (i.e., how they get along with parents, friends, teachers), school (i.e., school work), and physical complaints (i.e., headaches, dental problems, etc). If one reason for identifying and measuring health indicators is to help guide prevention efforts, and not merely to serve as a barometer, than more will need to be known about how adolescent's perceive risk and health.

CRITERIA FOR INCLUSION

Based upon the previous discussion, several criteria have been chosen to direct the selection of adolescent health indicators:

1. The indicators must focus directly on the adolescent, not on indirect enabling or disabling factors of the family or community. Although these other factors provide important clues to better understand causality and to direct research, with a limited number of indicators that can ben chosen it is more important to assess the adolescent directly.

2. The indicators must be justifiable according to either the degree of burden of suffering experienced by adolescents or else their economic burden to society. Indicators should focus on conditions amenable to either primary or secondary preventive interventions. With a limited number of indicators that can be tracked over time, care should be taken to chose only those measures that, if improved, will produce the most good for the most people.

3. The indicators must be measurable and, to have the greatest impact, must be easily understood by society. Meeting this criteria will be tricky. The tendency will be to chose indicators that are simple and universally measured on a routine basis. Because of the complexity of issues involved, there are no

clear markers of adolescent health that are as easily followed as those for pregnancy and infancy. The ideal situation would be to measure adolescent indicators annually because of the rapid and substantial psychosocial changes youth experience. In reality, there will need to be a compromise between choosing health measures that are relevant and choosing measures that are assessed by existing health surveys.

4. The indicators must be amenable to reporting by various distinctions that are consistent developmentally. As a minimum, these should include age, gender, race/ethnic group, and preferably, family characteristics. Care should be taken to ensure that the package of indicators are balanced and include health promoting factors as well as markers of health problems.

As a basis for the justification of health indicators, the conceptual framework developed by the Public Health Service (PHS) in its document, Healthy People 2000: National Health Promotion and Disease Prevention Objectives was used (7). In this report, the PHS identified 298 health objectives in 22 separate priority areas. The purpose of having health objectives is to guide public research, education, and services toward reducing preventable death, disease, and disability. Approximately 70 of these objectives related directly to adolescents and have been published by the AMA (8).

The PHS Year 2000 objectives are divided into three groups, those that address health status, those that address risk reduction and health promotion, and those that address health services. *Health status* measures relate to current disease, death, or disability; *risk reduction* indicators relate to reducing the prevalence of risks to health or to increase behaviors known to reduce such risks; and *service* indicators are relate to increase comprehensiveness, accessibility, and/or quality of preventive services and preventive interventions.

The three categories described above were used to organize possible adolescent health indicators. This distinction serves to both organize the health objectives and to promote integration of efforts among federal and private health initiatives that might use health indicator data.

Once criteria and organizational structure were identified, the next step in identifying a set of adolescent health indicators was to use current epidemiological data and data on health services to identify a list of possible measures that could be included in each category (see Tables 1-3). This list was compiled by reviewing existing papers and source books that describe markers of adolescent health and well-being. The most commonly used markers were included in the list.

INSERT TABLES 1-3 HERE

Next, a group of national experts was asked to rank order the markers in each of the three categories as to how useful each was as a health indicator. Experts were chosen who represented a range of professional disciplines.[1] The average rank order for each category was computed. Indicators that were closely aligned were collapsed to produce the final listing (see Table 4).

INSERT TABLE 4 HERE)

CONCLUSION

A paradigm based upon the PHS Year 2000 Health Objectives was used to select groups of indicators for tracking adolescent health. This approach produces three types of indicators: health status measures, risk reduction and health promotion measures, and health services. Based upon the rankings of a national group of experts, a small number of indicators was selected for each of the three categories. The recommended health indicators for adolescents are:

---

[1] The Expert Panel consisted of Charles Irwin, M.D., Anne Petersen, Ph.D., Barbara Ritchen, R.N., M.S.N., John Schowalter, M.D., Barbara Starfield, M.D., and Laurie Zabin, Ph.D.

> Health Status: The number of teens seen in emergency rooms with an intentional or unintentional injury.
>
> Risk Reduction and Health Promotion:
> The rate of teens who drink alcohol daily;
> The rate of teens who drove a motor vehicle after drinking during the past month;
> The rate of teens who carry a weapon to school.
>
> Health Services.
> The rate of teens with completed immunizations;
> The rate of teens who have a primary health care provider.

These indicators emphasize the importance of violence and injury to the health and well-being of adolescents and to society. They also underscore the causative role of alcohol in adolescent morbidity. Completed immunizations and having a primary health care provider are rather straightforward and traditional health service indicators that have inherent validity.

For the most part, these six indicators are already monitored on a regular basis currently. The number of teens seen in emergency rooms for injury is measured by the National Ambulatory Medical Care Survey (9). This annual survey, which was first done in 1992, includes data from a national probability sample of emergency rooms. The risk reduction and health promotion indicators can be obtained from the YRBSS and the Monitoring the Future Survey (3,4). The health service indicators can be obtained from the National Medical Care Utilization and Expenditure Survey and the National Health Interview Survey (10,11). Taken together, these indicators produce a well-rounded picture of adolescent health and well-being.

Table 1: Health Status Indicators

1. Rate of teens who are obese
2. Rate of teens who diet frequently
3. Rate teens who have iron deficiency anemia
4. Rate of teens with genital gonorrhea infections
5. Rate of teens who have had a pregnancy
6. Rate of teens seen in emergency rooms with a self-inflicted injury or overdose
7. Number of teens seen in emergency rooms with alcohol or drug related injury
8. Rate of teens who die from an alcohol-related motor vehicle crash
9. Days missed from school/work during the past year
10. Rate of teens with a chronic condition that results in some loss of ability to conduct normal physical, social, or recreational activities
11. Days hospitalized during the past year for conditions with preventable relapses, such as asthma and diabetes mellitus
12. Mortality rate, broken down by cause of death
13. Rate of victimization of violent crime
14. Percent of teens treated for emotional or behavioral problems in the past 12 months
15. Percent of teens who had an accident, injury, or poisoning in the past 12 months
16. Percent of teens with indicators of anxiety or depression

Table 2: Risk Reduction and Health Promotion Indicators

1. Rate of teens who smoke daily
2. Rate of teens who drank alcohol during the past month
3. Rate of teens who drink alcohol daily
4. Rate of teens who drove a motor vehicle after drinking during the past month
5. Rate of teens who disapprove of tobacco, alcohol, and drug use
6. Arrest rates for alcohol or drug related violations
7. Rate of illicit substance use during the past month
8. Rate of teens between who have had sexual intercourse
9. Rate of teens who used a condom at last intercourse
10. Rate of teens who carry a weapon to school
11. Rate of teens who participate in daily school physical education
12. Rate of teens who consume three or more servings daily of foods rich in calcium
13. Rate of teens who have at least one meal a day with their parent
14. Rate of teens who have discussed AIDS with their parents
15. Rate of teens who participate in an extracurricular activity
16. Rate of teens who value sexual restraint

Table 3: Health Service Indicators

1. Rate of teens with completed immunization (Dt booster, second MMR, HBV vaccine)
2. Rate of teens who had a routine (preventive service) visit in the last year
3. Rate of teens who had a dental exam during the past year
4. Rate of teens who have a primary health care provider or a clinic that serves as a "health care home"
5. Rate of teens who know that they can receive confidential health services related to reproductive health; physical or sexual abuse; and alcohol and drug problems
6. Rate of sexually teens who had pelvic exam (females) or genital exam (males) during the past year
8. Rate of teens who used psychological services during the past year
9. Rate of teens who are screened about sexual behavior
10. Rate of teens who are screened about use of tobacco products
11. Rate of teens who are screened about use of alcohol and other drugs
11. Rate of teens who are covered by either public or private health insurance

Table 4: Top Rankings by Category of Indicators

Ranking	Health Status
#1	Number of teens seen in ER with an intention or unintentional injury
#2	Mortality rate, broken down by cause of death, including deaths from alcohol-related motor vehicle crash
#3	Rate of teens who have had a pregnancy

Ranking	Risk Reduction and Health Promotion
#1 (tie)	Rate of teens who drink alcohol daily. Rate of teens who drove a motor vehicle after drinking during the past month. Rate of teens who carry a weapon to school
#2 (tie)	Rate of teens who smoke daily Rate of teens who had unprotected sexual intercourse at last episode

Ranking	Service
#1 (tie)	Rate of teens with completed immunizations Rate of teens who have a primary health care provider
#2	Rate of teens who have had a preventive service visit during which time they were screened for sexual behavior, use of tobacco products, and use of alcohol and other drugs.

REFERENCES

1. Marks, JS, Hogelin, GC, Gentry EM, Jones JT, Gaines KL, Lorman MR, Trowbridge FL. State-specific prevalence estimates of behavioral risk factors. Am J Prev Med 1985; 1:1-8.

2. Researching the Family: A Guide to Survey and Statistical Data on Families. Zill N, Daly M (eds). Child Trends, Inc: Washington, DC, 1993.

3. Kann L, Warren W, Collins JL, Kolbe LJ. Results from the national school-based 1991 Youth Risk Behavior Survey and progress toward achieving related health objectives for the national. Pub Health Reports 1993; 108 (suppl) 1:47-67.

4. Johnson LD, M'Malley PM, Bachman JG, Schulenberg J (1993). The aims, objectives, and rationale of the Monitoring the Future study. Monitoring the Future Occasional Paper No 34. Ann Arbor, MI).

5. Gans JG, Blyth DA, Elster AB, Gaveres LL. America's Adolescents: How Healthy Are They? American Medical Association: Chicago, IL 1990.

6. Millstein SG. A view of health from the adolescent's perspective. In Promoting the Health of Adolescents: New Dimensions for the Twenty-First Century. Millstein SG, Petersen AC, Nightingale EO (eds). Oxford University Press: New York, 1993, pp 97-118.

7. Healthy People 2000: National Health Promotion and Disease Prevention Objectives. U.S. Public Health Service: Washington, DC, 1990.

8. Healthy Youth 2000: National Health Promotion and Disease Prevention Objectives for Adolescents. American Medical Association: Chicago, IL, 1991.

9. Nelson CR, Stussman BJ. Alcohol-related and drug-related visits to hospital emergency departments: 1992 National Hospital Ambulatory Medical Care Survey. Advance Data 251, 1994.

10. Bloom B. Health insurance and medical care: Health of our nation's children, United States, 1988. Advance Data 188, 1990.

11. Butler JA, Winter WD, Singer JD, Wenger M. Medical care use and expenditure among children and youth in the United States: Analysis of a national probability sample. Pediatrics 1985; 76:495-507.

Indicators for Infant, Child, Preadolescent and Adolescent Health

Discussion

Prepared by Michael D. Resnick, Ph.D.
Associate Professor of Public Health and Pediatrics,
Director of Research,
Division of General Pediatrics and Adolescent Health,
Director, National Adolescent Health Resource Center
University of Minnesota

November 17, 1994

I am struck by the assemblage of talent in this room as well as the weightiness of the task of developing a meaningful set of social indicators for the population. The task, I believe, is to develop the tools that permit the painting of a portrait of America. We are here to assess the utility of different colors and brushes recommended by our topical authors. I am also mindful that while at this point in the evolution of scientific knowledge and method we are compelled by the rhetoric of numerality, there is behind the statistical portraits many faces, and the complicated contexts of their lives.

The users of this set of indicators will be formidably diverse - advocates, detractors, royalty and revolutionaries. Policy makers and pundits will use and misuse the fruits of these labors. Our most important constituents are those who seek through honest and concerted action, to promote and preserve the health and well-being of the people - all the people. We must feed their actions with information that is practical enough to be of utility, and also imaginative enough to capture something essential about the human experience and endeavor.

I imagine two contrasting points: the biologist who reduces us as living creatures to seven essential indicators: those of excretion, growth, irritability, locomotion, nutrition, reproduction and respiration.. On the other hand is the work of Stephen Boyden, the professor of human ecology at the Australian National University who wrote Western Civilization in Biological Perspective. In that work, he described the psychosocial indicators of life that are conducive to good health. He based his assessment on what was provided by hunter-gatherer societies, the social form in which Homo Sapiens have spent most of their evolutionary history. He suggests that this set of social indicators provides clues as to the universal health needs of the human species. These include an environment and lifestyle that provide a sense of personal involvement, belonging, responsibility, challenge, satisfaction, comradeship and love, pleasure, confidence and security.

What is clear is that to a growing number of observers, post-modern life no longer offers these qualities. And the growing absence of these elements in the lives of young people in particular is something that undermines their resilience and capacity to cope with the personal difficulties and hardships of everyday life.

Perhaps what we are looking for the most, is a set of social indicators that not only monitor trends in dying, distress, disability and discomfort, but indicators of sparkle, satisfaction and well-being - plus the elements that contribute to the uplifting or the stifling of the human spirit.

We have an excellent surveillance and monitoring system for human health when it comes to outbreaks, infection, and poisoning (including tainted ice-cream from my home state of Minnesota). But despite our increased knowledge about psychosocial etiologies of threats to the health of our young, we lack any kind of coherent, early response mechanism. We mobilize disaster relief when there are thirty deaths from a flood or tornado. But damage due to poverty, despair, hopelessness, or violence is declared a function of individual choice without attention to the influencing context within which it occurs.

My hope for this set of social indicators for infants, children, pre-teens, and adolescents is that they will function like the DEW line of the 1950's: the Distant Early Warning system against incoming threats to the nation. Our DEW line of the 1990's and beyond needs to reflect our deep understanding of contemporary threats to the well-being of young people, and a commitment to respond vigorously when public health disaster or sociological emergency is evident. I wish us well in this most urgent task.

The paper on prenatal and infant health indicators is by Paula Lantz of University of Michigan's School of Public Health and Melissa Partin of the Minnesota Department of Health. This is a careful consideration of a wide range of sources of indicators for our pre-born and youngest people. Their focus is on measures of infant mortality which is divided into neonatal and post-neonatal mortality rates; low birth weight; and prenatal care utilization. Our major sources are vital registration, medical records, and survey research data.

Vital records data almost completely cover births and deaths, data collection methods and forms are similar across regions, although some specific data elements such as gestational age, obstetric complications, medical interventions during pregnancy and substance use during pregnancy are often inaccurate or incomplete.

Data quality concerns are highest about information on causes of death, and race identification on death as well as birth certificates.

Medical records data are a powerful source of information not available through vital records or survey self-report, although procurement of such data can be difficult, as is the matching of prenatal care, maternal and infant hospital data, and subsequent health data for both.

Surveys offer a rich array of variables not captured in either previous source, but consistency, validity, and variations in measurement are often problematic.

The authors recommend that national and state infant mortality rates be produced by race or ethnicity, and maternal age. Cause-specific infant mortality rates should be produced annually, with analyses over time identifying trends in levels and patterns of infant mortality.

Low and very low birthweight data are continuously derived from vital records. Periodic data sources include medical records matched with samples of births; social surveys such as NSFG and NLSY and others.

Birthweight data suffer from under-reporting of extremely low birthweights (i.e. under 500 grams) and reporting bias linked to maternal social characteristics Better data on gestational age are needed to allow best possible use of birthweight and outcomes information.

Prenatal care data are also continuously derived from vital records, with periodic information available from other major national surveys which permit deeper level analyses. Measures of prenatal care have improved in documentation of the timing and quantity of prenatal care obtained by mothers, but little attention has been paid to the distribution and content of those visits. Problems of respondent self-selection and measurement of gestational age persist. Indeed, of the three indicators discussed: infant mortality, low birthweight and prenatal care, it is the latter that is most in need of overall improvement.

The authors describe other indicators useful in the assessment of prenatal and infant well-being, including fetal mortality rates, maternal health during pregnancy, APGAR scores, congenital anomalies, infant morbidity, and measures of growth and development. For all indicators, Lantz and Partin emphasis the fundamental importance of the translation of this information into best programs, policies and practices that promote health and well-being. Indeed, this is a continuous theme throughout all of these papers on social indicators for the young.

Barbara Wolfe and James Sears, both of the University of Wisconsin-Madison write on health indicators for pre-school children, ages 1-4. Their discussion is framed in terms of child poverty rates, and the national shame of non-coverage of many children by either public or private insurance. They pose three criteria for assessing the utility of current health indicators, including variability, validity, and reliability. Common indicators include global health status, medical care utilization, impairments, and other medical conditions.

Health status includes overall rating of health, description of engagement in usual activities - which needs to be linked to reasons for those limitations, and anthropometric measures. Medical care utilization typically queries on service use over a specified time frame, and typical sources of care. Insurance coverage which should include extent, nature and stability of coverage, and immunization history, particularly from administrative data rather than parental report are potentially useful measures.

Impairments include sensorimotor and developmental issues. Measurement of acute and chronic conditions is limited by issues of recall, variations in severity, and under-reporting when there is no provider contact.

The authors note the importance of environmental or contextual factors which is refreshing to hear from a department of economics. They focus on child care and measures of safety and unintentional injury both of which are relevant to health assessment, and under-scrutinized in most data collection.

For all suggested indicators, the analyses with greatest utility for action and intervention are those that disaggregate results by income, race and geography. In this lies the greatest potential for the targeting of programs and policies.

Barbara Starfield of the Johns Hopkins University discusses health indicators for preadolescent children. She frames her discussion in terms of specific health indicators, and health status profiles which are aggregated across multiple domains. Indicators such as mortality, morbidity, years of life lost are widely available, internationally comparable by and large, but often translate into categorical programs that address specific subparts of much greater, interrelated issues. In the words of writer H.L. Mencken, for every complex and difficult problem, there is always an answer that is simple, easy and wrong. Analysis of health profiles can escape this problem, but instrumentation for such assessment is only in its childhood or adolescence as far as psychometric assessment is concerned. Problems of comparability across instruments are commonplace, and in the case of some of these instruments, behavioral dependent variables abound with a stunning absence of independent variables. (One health services researcher was heard to lament: "What, questions don't have answers anymore??")

Starfield reviews the array of sources of health indicators, including vital statistics, hospital discharge data, interview, and clinical examination data, the latter of which may include laboratory test results. She notes poor reliability of examination data, including when conducted by physicians where there is only about 15% concordance between two physician raters on the presence or absence of abnormalities. This fact caused me to temper my disbelief of Saxon Graham's old epidemiological finding of 25 years ago that 33% of men misreport whether they are circumcised. By Starfield's data they are doing pretty well.

Do look at her Table 5: it utilizes thirteen criteria for assessing the utility of measures of service utilization, death rates, activity restrictions, and indicators of health conditions. On the basis of this, she recommends these indicators for preadolescents:

%death rates in total, and dichotomized as deaths from accidents and injuries, vs. medical causes.

%activity limitations, both total and by morbidity burden.

%hospitalization for conditions sensitive to primary care, and

%indicator conditions including communicable diseases, iron-deficiency anemia and elevated blood lead levels, and morbidity burden. A few words on the latter point. Aggregated measures of morbidity, the extent and severity of morbidity are critical informational elements as we move toward systems of care that are paid or reimbursed on the basis of their assessed ability to maintain and promote the health status of a defined population. In programs in health administration such as the one I teach in at University of Minnesota, training in epidemiological methods and measurement is predicated on the assumption that Integrated Service Networks or other large scale corporatized entities will be held accountable for quantified measures of the health status and functional effectiveness of their clients. Indices of diagnosed morbidity will prove to be an important component of these information systems. Important also will be Starfield's last recommendations, which are measures of health-influencing behaviors or conditions, such as availability of loaded firearms in the homes, and extent of television watching. Lest any of you raise an eyebrow at the idea of TV-watching as an early warning indicator, remember that we are first, absolutely first among ten industrialized nations in the percentage of thirteen year olds who watch five or more hours of TV per day. By the time a child reaches six years of age, on average they will have already seen more television than they will spend talking with their father for the rest of their life. We also know that one of the best predictors of academic performance in the 4th of 5th grade is whether the TV was on during family mealtimes during the preschool and kindergarten years. (En guarde! apologists for the media)

All of these indicators should be collected with the intent of analyzing and comparing findings by gender, race, geography, and social class status.

Art Elster of the American Medical Association's Department of Adolescent Health writes on health indicators for adolescents. He immediately sets forth as a criterion for health indicators their utility for purposes of advocacy, program, policy, and practice change. In the adolescent health arena, most reported indicators focus on health risk behaviors, with far less attention to social context, pro-social and health promoting behaviors - evidenced by the CDC's Youth Risk Behavior Survey, and Monitoring the Future from the University of Michigan.

This skew in our attention perpetuates a view of adolescence as riddled with problemness; it invites exaggeration of the perceived prevalence of distressed and destructive young people, and may inadvertently promote fear and hostility toward adolescents in general. Narrowness of perspective also invites sensationalism and over- focus on segregated problems rather than the adolescent as a purposeful actor living in the context of family and community.

Elster's conceptual frame for proposing needed health indicators is that threats to adolescent health emanate from the complex interplay of biological, psychological and sociological factors, with a growing eminence of social etiologies. Experimentation is normative, and normative definitions vary across communities and contexts. Adolescents also view the concepts of health risks and risky behaviors differently than adults, so moreso than with any other age group, the meaning of behaviors is critical. (Parenthetically, I applaud NICHD for its RFA on the meaning of unintended pregnancy, because it will engender research that helps us know what young people themselves think about sexual intercourse, intimacy, contraceptive use, and pregnancy.)

Elster argues for the selection of adolescent health indicators that are modifiable and amenable to action, understandable by those making decisions with and on behalf of adolescents, and analyzable along developmental lines, avoiding the lumping of, for example, 15 to 19 year olds which homogenizes middle and late adolescence. Remember, the distinctiveness of being nineteen years old is that this is the only year that you can live your philosophy of life: before 19 you haven't developed one, and after 19, you have to compromise too much with reality. Elster also reminds us that analysis of discrete, single health risk behaviors tends to obscure the co-occurrence of many behaviors, and their meaningful associations with relevant indicators of gender, class, culture, and functional effectiveness in other areas of adolescents' lives such as education, family and peer relationships, and work.

The recommended domains for assessment include health status, risk reduction, and health promotion measures. These would include, for example, measures of emergency room utilization due to injury, mortality by cause of death including alcohol-related fatalities, pregnancy rates, injury prone behaviors, tobacco use, unprotected intercourse, preventive service utilization.

Conclusion:

My optimism about the goals and process of this conference is that with a lot of effort and a bit of luck, the recommendations of this group may translate into actual data collection, and the availability of information that reflects the needs of young and very young people. I envision the means to paint a portrait of infants, children and youth that more closely reflect their physical, social, and economic realities at millennium. Our real goal is the statistical metaphor of Goodness of Fit: we want to maximize the fit between the health needs of these populations, and the indicators we collect that reflect those needs, their causes, and solutions. I have no doubt that large scale health care entities in the context of managed competition will look at these indicators and evaluate their utility for monitoring and assessing the health of populations for which they are accountable. We will help them to understand that as we move on the continuum of prenatality to infancy, to childhood, preadolescence and adolescence, we need to increasingly incorporate measures of the objective and subjective social environment, because of intimate connectedness with health status, well-being, health promotion, health demotion, or destruction.

Creating the tools that paint the portrait also shape the agenda for advocates and policy makers. There can be much ripple effect from our efforts here, as well there should be. I want to conclude with a beautiful story, as told by Elie Wiesel. It is a parable that teaches that whatever the question is, the answer is always in your domain. Once upon a time there was an emperor, and the emperor heard of a very wise woman. The wise woman was known for her powers. She knew how to listen to the wind, and interpret its melody. She knew how to describe the symphony of the stars. She understood the language of the birds. She knew everything. So the emperor said "Get her." They brought the wise woman to the emperor.

Emperor to the wise woman:

> "Is it true that you understand the language of the birds?"
> "I think so."
> "Is it true you know how to read the traces the wind leaves on the sea?"
> "I think so."
> "Is it true you know the symphony of the stars?"
> "I think so."
> "In that case," says the emperor, "I also heard that you know how to read someone else's mind. Can you read my mind?"
> "I think so."
> "In that case," says the emperor, "I have in my hands behind my back, a bird. Tell me: is it living...or dead?"
> And the wise woman was afraid. Maybe the bird was still living, and then the emperor - in order to prove a point - would kill the bird. So she waited for a very long moment, and then smiled, looked straight into the eyes of the emperor and said,
> "Majesty, the answer is in your hands."

Thank you.

Michael D. Resnick, Ph.D.
Associate Professor of Public Health and Pediatrics
Director, National Adolescent Health Resource Center
University of Minnesota
tel: (612) 624-9111
Fax: (612) 624-3972

"When the sentimentalist and the moralist fail, they will have as a last resource to call in the aid of the economist."--Sir Edwin Chadwick

EDUCATION

Indicators for School Readiness, Schooling, and Child Care in Early to Middle Childhood

Deborah Phillips
Board on Children and Families
National Research Council/Institute of Medicine

John Love
Mathematica Policy Research, Inc.

Paper prepared for conference on "Indicators of Children's Well-Being", November 17-18, 1994, Washington, D.C. Sponsored by the Institute for Research on Poverty, the Office of the Assistant Secretary for Planning and Evaluation (DHHS), the National Institute of Child Health and Human Development, and Child Trends, Inc. The opinions expressed in this paper are those of the authors only. They do not reflect the views of our home institutions. The authors wish to thank Richard Murnane and Don Hernandez for their insightful comments on an earlier draft of the paper.

INTRODUCTION

A national consensus has recently re-emerged regarding the importance of education, fueled in part by a perception that our schools are not doing an adequate job of preparing an educated citizenry for the 21st century. At the same time, national attention has been riveted on notions of outcome accountability for a variety of reasons ranging from frustration with the regulation of inputs to hopes that a reliable accountability system might provide persuasive evidence of the effectiveness of interventions for children and their families (Schorr, 1994). As a result, indicators that assess and track the school readiness and schooling of our nation's children are likely to become a particularly salient component of any effort to construct national indicators for children. Indeed, they will likely be used not only to track children's well-being, but also to assess the success or failure of our recent national experiment in school reform. A recent report from the Department of Education, Education Counts: An Indicator System to Monitor the Nation's Educational Health (NHES, 1991) states that "if the broad reform movement is to succeed, the United States must develop a comprehensive educational indicator information system" (p. 6).

The development of this system is beyond the purview of this paper. Indeed, its indispensability for a successful school reform effort is highly questionable. Indicators, in general, seldom offer appropriate tools for purposes of evaluation. On the other hand, an accepted and valid set of indicators can be a highly effective device for public communication and a significant lever for change. As such, efforts to construct a set of school readiness indicators that expand the richness, depth, and rigor of our understanding of children's well-being, and enables us to chart their educational progress from the child care through the middle school years, warrant substantial attention.

OUR APPROACH TO IDENTIFYING INDICATORS

A disclaimer is in order at the outset. Our training as developmental psychologists and our experience with program evaluation have prepared us to capture the contexts and complexity of children's lives, to search for explanations of the trends that characterize these lives, and to mistrust data that get far removed from the observational methods that our fields have labored long to develop and refine. Indicators, in contrast, emphasize simplicity, are designed to monitor rather than to understand children's development, and, by design, do not rely on labor intensive data collection methods.

We have, as a result, adopted an approach to the task of identifying indicators that draws upon our conceptual understanding of what to measure and then considers how best to quantify these concepts in the form of indicators. Specifically, we draw upon our knowledge of the developmental and evaluation literatures to identify dimensions of family and child well-being relevant to child care and early schooling that are most predictive of positive child outcomes in the short and long-term. We then discuss the implications of this empirical evidence for indicator data. In effect, we start with the goal of developing a set of indicators that measures the "right things", as noted by Brandon (1992).

In some instances, this approach points to a critical facet of development, such as "approaches to learning," for which no reliable indicator-type data sources presently exist. We hope, however, that our conceptual starting point will guard against the temptation to identify straightforward, easy to collect indicators that may be useless for policy purposes, or even misleading. We are particularly concerned about the tendency, over time, for indicators to take on a life of their own; to reify--rather than simply to reflect--the important parameters of child and family well-being. The strength of indicators is that they focus attention on critical issues. But, if we focus attention on the wrong issues, or on unreliable sources of information about the right issues, then we run the risk of misdirecting both public attention and public policy.

Consider the assessment of child care quality--the aspect of child care that is most strongly predictive of children's well-being (in contrast to use, type, or duration of child care). Several large surveys (e.g., National Household Education Survey, National Longitudinal Survey of Youth) have asked parents to report on quality features of their child care arrangements (e.g., staff-child ratios, total numbers of children, and staff qualifications). The reliability and validity of these reports, particularly for group care arrangements, are entirely unproven, and, for retrospective reports, are most likely very poor. Rather than propose an indicator based on parent reports of quality, we suggest searching for other child care indicators that can be accurately assessed.

In the domain of educational outcomes, we face a special challenge posed by the strong association in the United States between educational achievement and the demographic characteristics of the families whose children are being assessed, particularly their levels of maternal education. As a result, indicators of academic achievement could easily allow a state or school district to proclaim that its particular brand of education reform is especially effective--or to be subjected to criticism as ineffectual--when changing community demographics, rather than improved educational programs, could account for aggregate improvements or declines in school performance indicators. We strongly recommend that any sub-national or longitudinal reporting of educational outcomes be accompanied by information about (or adjustments for) the socio-economic status and ethnic composition of the population under consideration to guard against such misattributions.

SCHOOL READINESS

The concept of school readiness is central to each of the sets of indicators that are addressed in this paper and thus offers an appropriate point of departure. We care about school readiness because, as a nation, we are becoming increasingly concerned about the fact that children enter kindergarten with widely differing levels of preparation and, therefore, differing levels of functioning (see Bradburn, 1994). This causes us, on the one hand, to look backwards at variation in the resources and experiences to which children are exposed prior to school entry. Child care is included in this paper because it is now perceived as an environment that, among other goals, should help to prepare children for school. We also look ahead to children's differential progress through the school system which is now viewed as a function, in part, of their uneven status at school entry. Thus, the middle childhood years are included because this is presumably a useful point at which to take stock of the adequacy with which we have prepared children for school.

Conceptualizing School Readiness

Although any one of these premises is open to debate (e.g., child care should be a place where children play, free of the pressures of being prepared for school), we have chosen not to delve into the intricacies of these controversies. We cannot, however, so quickly by-pass the controversy that has surrounded the current state of knowledge and debate about the concept of school readiness itself. In practice, the selection of indicators involves the selection of social goals. Moreover, because of the political significance of social indicators, we are appropriately cautioned to assure that they are accepted and readily understood by the public (Moore, 1994).

Efforts to conceptualize school readiness, while widespread, are in their infancy and characterized by controversy. Two important tensions, with relevance to constructing indicators, are particularly prominent. The first concerns the distinction between school readiness and learning readiness. School readiness is generally approached as a school entry measure--a fixed standard of development sufficient to enable children to fulfill school requirements and to absorb the curriculum content (Kagan, 1994). This stands in contrast to concepts of learning readiness that acknowledge the fluid and cumulative nature of development, and typically adopt a more idiosyncratic, than normative, perspective. This is possible, in part, because concepts of learning readiness are not tied to a specific set of institutional requirements or expectations. Indeed, some assert that all children are born ready to learn even though not all are ready for school.

The second tension exists between the prevailing emphasis on children's readiness for school (the child outcome focus) and the relative inattention that is presently being paid to the extent to which schools are ready for the children they are now receiving and responsible for educating (the institutional focus). This tension derives from the concerns of many that assessments of young children's readiness will be misused to "blame" children and their families for low levels of early learning when, in fact, at least a portion of responsibility should lie with schools that vary in the extent to which they are receptive places for young children with differing characteristics and backgrounds (see Love, Aber, & Brooks-Gunn, 1994). Stated more constructively, efforts to promote the early success of children in school surely entail offering children beneficial early inputs and experiences (ranging from good nutrition to good books) and assuring that the classrooms and teachers they first encounter are receptive and affirming of their backgrounds, capabilities, and interests.

For the task at hand, we have been asked to focus on indicators that pertain directly to child outcomes and children's well-being (and to avoid indicators of institutional or jurisdictional performance). We strongly recommend, however, that a comprehensive effort to develop childhood indicators include indicators of schools' readiness for the diverse populations of young children they must now educate.

Measuring School Readiness

The status of efforts to develop measures of school readiness is rudimentary, at best. And, they too, are immersed in controversies such as the appropriateness of such assessments for language minority children, and their role in determining school entry and tracking for very young children. This is murky and value-laden territory.

Yet, charged by the President and the 50 state governors in 1990 to assure that "by the year 2000 all children in American will start school ready to learn" (a goal that was lent the weight of law with the recent passage of the Goals 2000 legislation), a number of states have been designing and implementing their own readiness assessment systems. At the national level, the National Center for Education Statistics is supporting the development of a new assessment of readiness through the Early Childhood Longitudinal Survey (ECLS) (West, 1992). One of the primarily rationales for this survey, which is projected to go into the field in 1998, is "the scarcity of data on children's preparation for school, their transition into school, and their progress through the primary and elementary grades" (Bradburn, 1994). Focusing primarily on children in kindergarten through fifth grade, the ECLS includes a cohort of Head Start children. Although, as a longitudinal survey, the ECLS will not provide an on-going source of indicator data, it does offer a rare opportunity to develop indicators of school readiness, early schooling and child care, including quality of care for center-based arrangements.

In addition, the Office of Educational Research and Improvement (OERI) in the U.S. Department of Education is being reorganized to better fulfill its mission, which includes monitoring the state of education. The new OERI is structured around five national research institutes, including the National Institute on Early Childhood Development and Education and the National Institute on Student Achievement, Curriculum, and Assessment (OERI, 1994). The domain of readiness, schooling, and child care indicators bears directly on the agendas of these new institutes. Given the importance of data from the Department of Education for the indicators that we discuss, some coordinated planning would be highly desirable.

Most recently, the second author and his colleagues (Love, Aber, & Brooks-Gunn, 1994) have proposed an assessment system to help schools, communities, and states determine how effectively they are supporting and promoting children's school readiness. This system is designed to be implemented by most school districts in the context of their kindergarten registration procedures. If fully implemented, it too would offer a rich source of indicator data at district, state, and national levels of aggregation.

Absent the ECLS and the assessment system proposed by Love et al., we must fall back on current conceptualizations of school readiness and adapt them to our present purposes. Most fundamentally, conceptions of school readiness acknowledge the vast amount of school-relevant learning that occurs long before formal instruction is introduced at school entry. Empirical documentation of the significance of early learning has focused on early literacy acquisition (Hakuta, K., & D'Andrea, 1992; Snow, 1983), but growing evidence is now revealing the importance of early experience for numerical knowledge, as well (Case & Griffin, 1990; Griffin, Case, & Sandieson, 1992; Siegler & Robinson, 1982). Beyond the acquisition of early concepts and knowledge (e.g., the alphabet and the number line), a large literature has documented the many ways in which children's home environments instill the behavioral and motivational repertoires that enable children to enter school eager and ready to learn (Entwisle & Alexander, 1990; Stipek, 1988). Accordingly, a central challenges is that of deciphering those aspects of children's pre-school experiences that will provide a valid portrait of their preparation for school.

Once a child enters school, the assessment of readiness has received more attention, compared to the pre-school period. Of particular relevance to our task is the work of the Goal 1 Technical Planning Group of the National Education Goals Panel (December, 1993). The Planning Group has identified five dimensions that encompass the wide range of abilities and experiences on which early learning and development depend. Each dimension includes a number of criteria for assessment. These are:

Physical well-being and motor development:

- *Physical development* (rate of growth and physical fitness)
- *Physical abilities* (gross and fine motor skills, oral motor skills, and functional performance)

Social and emotional development:

- *Emotional development* (feeling states regarding self and others, including self-concept; the emotions of joy, fear, anger, grief, and so forth; and the ability to express feelings appropriately, including empathy and sensitivity to the feelings of others)
- *Social development* (cooperation, understanding the rights of others, ability to treat others equitably, ability to distinguish between incidental and intentional actions, willingness to give and receive support, ability to balance one's own needs with those of others, creating opportunities for affection and companionship, and ability to solicit and listen to other's points of view)

Approaches toward learning:

- *Predispositions* (gender, temperament, cultural patterns and values)
- *Learning styles* (openness to and curiosity about new tasks and challenges, task persistence and attentiveness, a tendency for reflection and interpretation, and imagination and invention)

Language usage:

- *Verbal language* (listening, speaking, social uses of language, vocabulary and meaning, questioning, creative uses of language)
- *Emerging literacy* (literature awareness, print awareness, story sense, and writing process)

Cognition and general knowledge:

- *Knowledge* (physical knowledge, logico-mathematical knowledge, and social-conventional knowledge)
- *Cognitive competencies* (representational thought problem solving, mathematical knowledge, and social knowledge)

In this paper, we narrow the lends to encompass the final three dimensions. See papers by Wolfe, Starfield, Aber, and Love (this volume) for discussion of the other dimensions.

Indicators of School Readiness

Drawing upon the National Education Goals Panel's (NEGP) conceptualization of readiness and the research literature on this topic, we suggest that indicators of school readiness focus on (1) Exposure to reading at home, (2) Exposure to pre-numeracy experiences at home, (3) Approaches to learning, (4) Emergent literacy and numeracy development, (5) Proportion of kindergartners deemed "unready" for kindergarten, (6) Parental attitudes and expectations, and (7) Access to some instruction in the child's native language. The home environment provides the focus for this section of the paper; children's child care environments are discussed in the next section. Table 1 presents the proposed list of school readiness indicators, distinguishing between those that are currently available and those that need to be developed.

Exposure to Reading at Home. Children's pre-literacy interactions in the home have been found repeatedly to differentiate children who are readily able to acquire age-appropriate information at school entry from those who are not. Specifically, a large and sophisticated literature has documented the predictive role that children's exposure to environments that are rich in discourse and literacy experiences plays in their reading levels at kindergarten and first grade (Dickinson & Beals, 1994; Goldenberg, 1987). The extent to which these experiences are provided to children is, in turn, affected by maternal education and parents' views about how children learn to read, write, and use numbers. Opportunities to acquire literacy skills at home, nevertheless, provide a highly valid proximal indicator of educationally significant early experiences.

Some of the most important aspects of these opportunities would be difficult to capture with indicators, including the extent to which parents depart from simply reading the text to engage children in conversations about the text and the extent to which children are encouraged to talk about past and future events. But, the number of books in the home, particularly the number of children's books, and parents' reports of time spent reading children's books to their children have been found to offer reasonable proxies for the home literacy environment.

Current indicators could be developed from the National Household Education Survey:93. This telephone survey of a representative sample of households with 3- to 7-year olds, sponsored by the Department of Education and first implemented in 1991, asked parents a series of questions about home activities that are relevant to the early reading environment. These include:

- whether the child pretends to read
- number of children's own books in the home
- frequency of reading to the child
- frequency of storytelling
- rules governing content and hours of children's television viewing (may bear on opportunities for reading experiences at home)

Factor analyses of data from the 1991 wave of the NHES, carried out by Zill and colleagues (Zill, Stief, & Coiro, 1992), identified four scales focusing on (1) activities with the child at home, (2) activities with the child outside the home, (3) educational materials in the home, and (4) rules about television viewing. The scales show good internal consistency and may offer an alternative to reliance on individual items.

This (or similar) information will be obtained in the NHES:95, and we understand that the NHES may be planning a parent component in the 1996 wave. Subsequent assessments will occur at 2-year intervals. The child well-being module of the SIPP (in the field) also asks parents of infants through 5-year olds about the frequency of reading to the child (ages 0 to 5) at home, and a set of questions about television viewing (rules about what shows, total hours and how early/late the child can watch). One note on this new module is in order: it is not clear whether this will be an on-going component of the SIPP. We would like to highlight the importance of repeating this module on a regular schedule.

Each wave of the mother-child module of the National Longitudinal Survey of Youth (NLSY) includes a modified version of the HOME Scale--a well validated and widely used assessment of the home learning environment. The mother-child supplement is a biennial survey, beginning in 1986, of the children of a nationally representative sample of women 14-21 when they were first interviewed in 1979. The children are assessed beginning at age 3 and interviewed directly beginning about age 10. These data are limited, however, by the basic design of the NLSY. Most notably, the older children are children of early childbearers and the younger children are children of later childbearers. The NLSY will eventually include children of older and younger childbearers at each age and, as such, will prove more useful as a possible source of indicator data.

Finally, we recommend that the development of the ECLS protocol be observed carefully as a potentially valuable "testing ground" for each of the readiness constructs that we have identified. We will re-emphasize this point only in those instances where we want to recommend that a particular construct, not presently highlighted in the plans for the ECLS, be seriously considered for inclusion in this study.

Exposure to Pre-numeracy Experiences. A parallel literature has focused on identifying the home experiences that distinguish children who come to school with an intuitive sense of numbers and how they work from those who do not. While not as well developed as the knowledge base about pre-literacy experiences, beneficial pre-numeracy experiences include board games and card games that involve numbers, as well as the engagement of children in conversations and other activities that associate number with quantity (e.g., sorting laundry or picking up toys) and teach children to think in terms of a mental number line. It is not simply the act of counting that matters, it is exposure to the functions and meaning of counting.

The challenge for an indicators project is one of identifying meaningful indicators from among the array of important experiences that have been identified. We are not aware of any current, representative data sources that inquire about pre-numeracy experiences, and propose that this be a priority for the development of new indicators data. An appropriate focus, parallel to indicators that capture books (resources) and reading (experiences), would be on the availability of counting games and toys (resources) and time spent playing with/explaining numbers to children as distinct from simply getting them to count to ten. These are admittedly far more complicated questions for parents to answer than those regarding literacy, and substantial work would be entailed in developing reliable and valid indicators. But, a growing literature on children's math achievement suggests that the effort would have large pay-offs.

The NHES:93 included a question regarding the frequency of playing with games and toys at home, as well as one inquiring about how high the child can count. Perhaps, in future waves, a probe about the types of games could be added and the query about counting could be replaced with a more meaningful item regarding numeracy experiences. The opportunity provided by the NHES to ask about pre-numeracy experiences in the context of other questions about parent-child activities in the preschool years is well worth exploring.

Approaches to Learning. As important for school achievement as children's early exposure to school-related concepts and skills is the early encouragement of their motivation to acquire and marshall this knowledge as they progress through school. Behaviors such as task persistence, impulse control and attentiveness are likely to improve children's adjustment to structured elementary school classrooms (Benasich, Brooks-Gunn, & McCormick, 1992; Lee, Brooks-Gunn, Schnur, & Liaw, 1990). The development of enhanced self-regulatory abilities (such as delay of gratification and impulse control) predicts academic competence (SAT scores) more than a decade later (Mischel, 1984). And, "personal maturity" in preschool, which includes a large self-regulatory component, predicts achievement in reading and math in elementary school (Entwisle, Alexander, Pallas, & Cadigan, 1987).

Children's approaches toward learning include curiosity, creativity, independence, cooperativeness, and persistence. This construct calls attention to the important distinction between children's repertoire of skills and knowledge, on the one hand, and their engagement in learning and self-concept as a learner, on the other hand. The Goal 1 Technical Planning Group identifies four components of "Approaches toward learning": (1) Openness to and curiosity about new tasks and challenges, (2) Task persistence and attentiveness, (3) A tendency for reflection and interpretation, and (4) Imagination and invention. This group further speculates that "approaches toward learning is the least understood dimension [of school readiness], the least researched, and perhaps the most important". We agree.

Other investigators have focused on somewhat different, but closely related, components of a child's approach to learning. Bronson (1994) has emphasized the "ability to carry out developmentally appropriate goal-oriented tasks in an independent, self-regulated manner". Component behaviors include selecting tasks appropriate to one's level of skill, organizing task-relevant materials, using effective task attack strategies, resisting distraction, trying repeatedly (persisting) when necessary, and, ultimately, completing tasks successfully" (p. 23). Bronson has developed a detailed observational measure that captures these constructs. Aber and his colleagues (Aber, Molnar, & Phillips, 1986) have used the term "disposition to learn" and emphasize the inseparability of cognitive from socio-emotional, motivational, and personality development. particularly during the preschool and early school years. Notions of self-regulatory behavior, as described above, are also featured prominently in this literature, with the preschool years identified as a particularly sensitive stage for their development (Aber et al., 1986).

Valid measurement of these constructs entails labor intensive methods: classroom observations of children or the administration of a set of child assessments. Such measures exist (e.g., Torrance's Thinking Creatively in Action and Movement measure for 3- to 8-year olds, 1981), but are not likely to be widely enough used to form the basis for a representative set of indicators. Teacher ratings can be used (e.g., Love et al., 1994, propose the self-control and cooperation subscales of the Social Skills Rating System [Gresham & Elliiott, 1990] for their assessment system), and offer a more practical source of indicator data. This is clearly a topic that warrants a high priority for the development of improved indicators.

We would like to make a particularly strong case for instrument development in this area in conjunction with the ECLS. This survey affords a rare opportunity to measure approaches toward learning, although inclusion of such assessment is not presently a priority. We believe that an investment of this sort now, given the timing of the ECLS, would reap substantial benefits for future efforts to track important indicators of school readiness.

Emergent literacy and numeracy development. These indicators would serve to capture, at school entry and during the early elementary years, the skills and knowledge in literacy and math that beneficial home pre-literacy and pre-numeracy experiences have been found foster. Language is central to learning in all domains of achievement, and is also the dimension of early learning that kindergarten teachers identified as the area where most "unready" children have difficulty (Boyer, 1991).

Measurement of literacy development is not straightforward. Ideally, it would encompass aspects of form (structure or syntax, including recognition of the alphabet), content (meaning or semantics; the ability to comprehend), and function (use of language to communicate; to acquire information)--each of which has its own developmental timetable. For our present purposes, two commonly accepted domains of emergency literacy require consideration: verbal language, including listening, speaking, social uses of language, and vocabulary and meaning; and literacy, including literature awareness, print awareness, story sense, and writing processes (see Goal 1 Technical Planning Group, 1993).

A special challenge in this area concerns children whose primary language is not English--a sizeable and growing share of the pre-school and elementary school population (see Phillips & Crowell, 1994). Whatever shape efforts to track literacy take, it will be critical to include immigrant and non-English-speaking children, as is the partially case with the Department of Education's Prospects study of Chapter I services (Puma et al, 1993), which includes Spanish-speaking children.

Numerical-mathematical knowledge is also heavily stressed in elementary school curricula. As with literacy skill, striking differences are found in the mathematical understandings that children bring to school (see, for example, Griffin, Case, & Siegler, 1992). A significant number of low-income children, for example, have been found to be unable to tell which of two numbers is bigger or smaller (e.g., 6 or 8) or which number (e.g., 6 or 2) is closer to 5. Yet, this is precisely the knowledge on which the solving of first-grade addition and subtraction problems is dependent. The concern here is that many children enter school without knowledge that their teachers assume they have, and are then left behind as their early school instruction departs from a baseline that they have never achieved.

Measurement of early number knowledge is at a more rudimentary stage than is the case with early literacy knowledge. The major challenge is that many children are able to count, but they do not have a sense of the "number line"--of how numbers relate to quantity and to sequencing--which is, in fact, the critical numeracy knowledge at school entry. A child's ability to count can actually camouflage the absence of adequate numerical knowledge.

At the time of kindergarten entry, there is sparse data from which to draw national indicators of literacy and numeracy knowledge. The NHES:93 (and presumably NHES:95 and/or the 1996 parent survey) asks parents about their children's knowledge of color names and the alphabet, about whether the child can write his/her first name, and how high the child can count. We are not confident of the validity of these data, and question whether counting per se is a useful indicator of numeracy knowledge. The child well-being module to the SIPP may contain some relevant items in the future.

A major assessment of the cognitive skills and abilities of children exposed to Chapter I services is being conducted by the Department of Education (Puma et al., 1993). This study, called Prospects: The Congressionally Mandated Study of Educational Growth and Opportunity, is following 30,000 students across the U.S. in grades 1, 3, and 7 for five years. It's purpose is to evaluate the long-term effects of exposure to Chapter I services. In addition, a subset of these students is being observed in classroom settings. At a minimum, the ECLS should examine the protocol for this study so that some parallel data are collected. The Prospects study may also be a current source of indicator data regarding literacy and numeracy skill, albeit for only a segment of the population. A real strength of this study is its inclusion of immigrant and non-English-speaking students who speak Spanish.

Again, one of the most promising prospects for improved indicators is the Early Childhood Longitudinal Study (ECLS). This will offer the opportunity to assess children's readiness in the Head Start and kindergarten cohorts. Love et al. (1994) propose use of the Early Screening Inventory (Meisels, Wiske, Henderson, Marsden, & Browning, 1988) as a source of information on expressive (verbal) language, verbal reasoning, and knowledge of colors, letters, numbers, and writing in their strategy for a district-level kindergarten entry assessment system. Vocabulary development (e.g., the PPVT) is also a useful correlate of early literacy development (see Cazden, Snow, & Heise-Baigorria, 1990). It would not be very difficult to design a set of useful items to assess early math knowledge, either incorporating or modifying assessments used in empirical research (see Griffin, Case & Siegler, 1992, for example).

We further suggest that state-level kindergarten assessment data be examined as a possible source of indicator data. Many states have developed a battery of kindergarten screening tests, some of which are highly regarded (see, for example, the nationally normed Tests of Early Math Ability, and analogous tests in reading, writing, and language, developed by nationally recognized researchers in each area). Many states, in addition, are engaged in efforts to construct their own assessments of school readiness.

Proportion of Kindergartners Deemed "Unready" for Kindergarten. A readiness indicator that has high face validity, but may be more a reflection of differing school practices than of child well-being, concerns the proportion of kindergarten-age children who are deemed "unready" for kindergarten. Some of these children are placed in transition kindergarten programs or are asked to repeat kindergarten; others are assigned to special education services in kindergarten.

The NHES:93 asks parents to report whether their child attended one or two years of kindergarten, and whether the child received any special help in school for reading, arithmetic, speech, a learning disability, or English as a second language. Relevant information will be available from the child well-being module of the SIPP, which asks parents of 6 to 11 year olds if their child has repeated a grade, including kindergarten.

The consistency and validity of state- and district-level data regarding the educational status of kindergartners also warrant careful attention. The School Archival Records Search (Walker, Block-Pedego, Todis, & Stevenson, 1991) offers a uniform system for obtaining information about children's school experiences from school records. Data collection includes information regarding school attendance, achievement, retention, in-school and outside referrals for academic or disciplinary causes, placements outside the regular classroom or for special services, and negative narrative comments.

Parental Attitudes and Expectations. Once children enter school, dimensions of parenting such as parental monitoring (Crouter, MacDermid, McHale & Perry-Jenkins, 1990; Dishion, 1990; Zill & Nord, 1994), positive mutual participation (Bradley, Caldwell, & Rock, 1988; Moorehouse, 1991), and parental involvement in the child's schooling (Alexander & Entwistle, 1988) become important predictors of children's motivation and performance in school. Parental expectations regarding their child's school performance are also correlated with schooling outcomes (Stipek, 1988). At very young ages, however, most parents (and children) hold high educational expectations, thereby generating only minimal variability.

It appears that the deployment of expectations, in the form of actual involvement (help with homework, taking the child to the library, getting to know teachers), is the more potent and discriminating indicator for young children (although the NHES:93 reveals that nearly three-quarters of students in the 3rd to 5th grade had parents who showed at least a moderate level of school involvement). The fact that the parent takes the time to get involved communicates to the child that s/he considers school important and is likely to indicate that the parent provides other forms of encouragement and support for learning outside of school (Zill & Nord, 1994).

The NHES:95 will ask parents who are using Head Start, a prekindergarten program or other group care program if they worked at the child's program in the last month. We are not aware of any other source of nationally representative data that reflect this construct at the pre-kindergarten or kindergarten age levels. Perhaps the ECLS, or the proposed Survey of Program Dynamics of the U.S. Bureau of the Census, will provide pertinent information.

Access to instruction in the child's native language. Estimates of the number of students in U.S. schools with limited English proficiency range from 2.3 million (U.S. Department of Education, 1992) to much higher (Stanford Working Group, 1993). The current influx of new immigrant groups, some of whom also have relatively high rates of birth, will fuel continued growth in the number of students who enter school with little or no English proficiency.

These trends pose new opportunities, but also serious challenges, to U.S. educational institutions, including the early childhood programs that lay the foundation for children's school experience and achievement (see Phillips & Crowell, 1994). In California, for example, a recent study of 400 child care centers revealed that only 4 percent enrolled children from a single racial group (Chang, 1993). Nationwide, estimates suggest that 20 percent of the children enrolled in Head Start speak a language other than English (Kagan & Garcia, 1991). In the D.C. public schools, over 100 languages are now represented.

Coinciding with these demographic trends, research now suggests that some degree of consistency in young children's exposure to their native language may be important for their later linguistic development and learning. Specifically, children younger than 5 years old are still acquiring the basic grammatical and phonological aspects of their first language. It appears that students can more readily become literate in a second language once literacy has been established in the home language (Snow, 1992). Moreover, if English is introduced at a very young age to a non-English speaking child, proficiency in the home language can be disrupted, with possible adverse consequences for the child's communication with parents and the home community.

For these reasons, we feel that it is extremely important to include consideration of language issues in any contemporary discussion of readiness, child care, and schooling indicators. The NHES:95 contains questions concerning the language spoken at home and the language of the child's caregiver/teacher. In addition, the Organization for Economic Co-operation and Development (OECD), with support from the National Center for Education Statistics, has recently published Education at a Glance, which summarizes 38 educational indicators from the OECD countries (OECD, 1993). Among these indicators is information on the percentage of children who say they usually speak the same language in school and at home. The information is based on a special survey conducted by the International Association for the Evaluation of Educational Achievement (IEA) and the Educational Testing Service (ETS), and includes only 9- to 14- year olds. We recommend that a down-age extension of this information be developed for future use in the child and family well-being indicators project.

Pertinent data could be obtained by adding questions about language of instruction and languages of students in the Schools and Staffing Surveys (school survey and teacher survey), conducted by the National Center for Education Statistics. This unified set of surveys profiles the nation's elementary and secondary school system, with the third administration conducted during the 1993-94 school year. The school survey includes information about student characteristics and about types of programs and services offered. The teacher survey collects data from teachers regarding their education, training, and teaching experience, among other things. The proposed Survey of Program Dynamics of the U.S. Bureau of the Census may also offer a source of information about languages used at school.

Priority Indicators

We propose three priority indicators. First, given strong evidence regarding the importance of early reading experiences for later success in school, and the growing policy interest in programs that promote these experiences (e.g., parent education, Early Head Start, Even Start), we include exposure to reading at home as a priority indicator. However, since exposure (differentiated from *where* the exposure occurs) appears to be the important variable, we encourage efforts to collect comparable data regarding pre-literacy experiences *across* the home and child care settings that children inhabit prior to school entry.

Second, few would dispute the importance of capturing indicators of children's earliest school performance in the areas of literacy and number knowledge. Early performance is a powerful predictor of later performance, and offers a useful proxy for the extent to which children are coming to school with the types of skill and knowledge that teachers typically expect and often assume as a point of departure for formal instruction.

Third, given the rapid diversification of the preschool population, substantial evidence regarding the importance *during the early years of language development* of support for the child's native language, and the availability of a data source (NHES:95), we recommend the inclusion of access to *some* instruction in the native language for children whose primary language is not English as a critical indicator of "readiness". This recommendation is not intended to detract from the importance of assuring that young children receive instruction in English--an aim that many non-English speaking parents appear to endorse for their children. Rather, we interpret the current literature to suggest that abrupt and discontinuous shifts from one language at home to another language at school may interfere with young children's first- and second-language development. This may become less important at later stages of schooling, although we would like to see some consideration given to bilingualism among older children in light of the diverse population and global economy in which today's generation of students will need to function productively.

Finally, we repeat our recommendation that the development of indicators of attitudes towards learning be a high priority for the future, with special attention paid to the opportunities that the ECLS provides along these lines. As we consider the future, we propose that access to educational technology at home and in preschool and kindergarten settings be added to the list of readiness indicators. We believe that this topic will rapidly become increasingly important for children's preparedness for school, as well as for considerations of equity of access to resources that facilitate success in school.

CHILD CARE

Children in the United States are negotiating the transition from home to school at younger ages than was true even a decade ago. Most children's initial exposure to a school-like setting used to occur when they entered kindergarten or first grade. Today, preschool and child care environments are playing this role in the lives of ever growing numbers of youngsters. As of 1990, 55 percent of low-income children aged three to five were enrolled in a school, child care center, or Head Start program (Brayfield, Deich, & Hofferth, 1993); 40

percent of all 3- and 4-year olds were in some form of group care or preschool program as of 1991 (Casper, Hawkins, & O'Connell, 1994). State and national welfare reform initiatives are likely to fuel substantial growth in these numbers, including growth in the number of infants and toddlers in non-maternal care settings.

At the same time, there has been growing recognition that the precursors of school success are found in the earliest years of life and that substantial learning occurs before children first encounter formal academic instruction. It is not surprising, then, that child care and preschool are no longer seen simply as a place where children play and have fun with their age-mates. Concerns about the educational attainment of the country's children have refocused attention on early childhood settings as places where children also get ready for school.

The educational significance of children's early care and preschool settings was prominently affirmed in 1990 when the President and the 50 state governors established the first of six national educational goals: "By the year 2000 all children in America will start school ready to learn". Assuring that "all children will have access to high quality and developmentally appropriate preschool programs that help prepare children for school" was identified as one of three objectives that accompanied the goal statement. In this context, it is critical that there be a close articulation between indicators of children's well-being across the preschool and school-age years.

Fortunately, there is a substantial research literature on the developmental consequences of child care from which to discern the "right things" to include in a national effort to assess and track children's well-being. Unfortunately, however, scant attention has been devoted to translating this empirical literature into a list of indicators. A preliminary effort was recently launched to develop a set of indicators of "improved results of the childhood care and education system" (Galinsky, personal communication, August 1994). But, to date, this initiative has not attempted to map its set of 24 proposed indicators onto available data sources. Further, most prior efforts to develop a set of education indicators curiously, though predictably, by-passed the pre-school years, with the notable exception of the post-1990 initiative of the NCES reported in Education Counts (NCES, 1991).

Indicators for Child Care

What aspects of child care warrant national attention in the context of an indicators project? Research evidence on child care indicates that the well-being of children depends primarily on the quality and continuity of their care settings and providers (Hayes, Palmer, & Zaslow, 1990). For low-income children, access to Head Start, school-sponsored prekindergarten programs, and other early intervention programs appears to be developmentally advantageous, at least in the short term. For school-age children, the absence of child care during hours when their parents cannot provide supervision is associated with adverse developmental outcomes, particularly for children under age 13 who live in urban areas (Long & Long, 1981; Coleman, Robinson, & Rowland, 1990; Vandell & Posner, in press). Parents' inclinations and ability to provide these features of care for their children, in turn, depend on issues of access/eligibility and of cost. Finally, recent evidence suggests that parents' perceptions that they are using arrangements that constitute their preferred choices affect their own efforts to attain economic self-sufficiency (Meyers, 1993). Thus, indicators of child well-being that relate to child care should focus on six general areas: (1) Quality of child care settings and providers, (2) Stability of care, (3) Access to early intervention programs on behalf of eligible populations, (4) Proportion of children under age 13 in latchkey situations, (5) Costs of care relative to family income, and (6) Parent choice. It is also worth noting that (3) (5) and (6) also raise critical issues regarding equity of access to decent child care options--a relatively neglected perspective on child care that is richly deserving of attention in this project.

Table 2 presents our proposed list of child care indicators. Two dimensions of care are conspicuously absent from our list of child care indicators: Use of child care and Type of care. Despite public anxiety regarding the dramatic shift from mother care to other care (and, specifically, to market care) that has characterized the last two decades, research has repeatedly demonstrated that the use/non-use of child care is not meaningfully associated with young children's development. Similarly, given the wide range of quality that characterizes every type of care, children's well-being does not appear to be differentially affected by the type of child care in which they are enrolled (e.g., center, family day care home, relative). As noted above, it is not whether or where children are being cared for that matters; it is how well.

Quality/Characteristics of Care. Quality of care is a heterogeneous construct, although most evidence suggests that "good things go together" in child care. We stress indicators of trained and educated staff, child:staff ratios and group size, staff salaries, and, for home-based settings, regulatory status and connection to provider networks. These are the quality variables, from among the large repertoire of quality indices that have been assessed, that have shown the most consistent and strongest associations with children's development in both center-based and home-based child care settings (Galinksy, Howes, Kontos, & Shinn, 1994; Hayes, Palmer, & Zaslow, 1990; Helburn, et al., 1995; Whitebook, Howes, & Phillips, 1989). Data sources that are worth exploring and developing in this regard include state regulatory data and a possible downward extension of the protocol that is being planned for the Early Childhood Longitudinal Survey.

There are no on-going sources of nationally representative data on child care that we feel confident recommending as a current source of indicators of child care quality. Although some national datasets include maternal reports of ratios and provider training, for example (NHES, NLSY), we are not confident of the validity or reliability of these data. Even self-reports of ratios from center directors, let alone mothers, have been found to be poorly associated with observational data (Phillips, Voran, Kisker, Howes, & Whitebook, 1994). Given the importance and likely weight that would be given to indicators of child care quality, we feel that it is important to wait for the development of reliable indices.

Stability of Care. Children who have experienced multiple changes in child care providers and arrangements prior to school entry show poorer developmental outcomes in both the short and long term (Cummings, 1980; Howes, 1988). It follows that one of the most important indicators of children's well-being in the context of child care cannot be captured with "snapshot" data. Rather, it is important to capture the patterning of care over the early childhood years.

These are difficult data to obtain given that their most reliable source is prospective accounts of children's concurrent and sequential child care arrangements. As a shortcut, however, we feel that it is worth obtaining mothers' counts of the total number of child care arrangements that they used for each child prior to kindergarten entry. Ideally, a variable that controls for the duration of children's reliance on child care, such as the average number of arrangements per year, would also be constructed. These indices are being collected prospectively in the NICHD Study of Early Child Care and data regarding their validity and predictive power will soon be available (see the NICHD Early Child Care Network, 1994).

The child care module of the SIPP will provide information about the total number of arrangements and changes in arrangements during the past 12 months. The pertinent questions are currently asked only of working mothers, but, in the next wave, will be expanded to include non-working mothers and non-work hours for all mothers. An additional improvement would involve asking specifically about the number of concurrent and sequential arrangements since the child was first placed in non-maternal child care. The NHES:95 includes questions that inquire about simultaneous care arrangements. Although parents may not be entirely accurate, we expect that the relative ranking of families using very few, a moderate number, and a high number of arrangements could be derived from parent reports.

Proportion of Eligible Children in Early Intervention Programs. We include this variable for two reasons. First, three major studies have now documented the higher quality of care that characterizes Head Start, Chapter I, and other school-sponsored early childhood programs when compared to community-based child care programs, particularly those that do not receive substantial public subsidies (Helburn et al., 1995; Layzer, Goodson, & Moss, 1993; Phillips, et al., 1994). Numerous studies examining the outcomes associated with early intervention programs, including Head Start, have documented their positive short-term (and sometimes long-term) effects on school achievement. Enrollment in these programs may, therefore, serve as a proxy for access to quality child care settings.

Second, given our national commitment to supporting several early intervention programs for low-income children (e.g., Head Start, Early Head Start, Chapter I pre-kindergarten, Even Start), it strikes us as highly appropriate to obtain estimates of the proportion of eligible children served.

It is a challenge, however, to obtain an accurate estimate of the proportion of eligible children in early intervention programs. The biggest problem concerns the fragmentation of data sources and the difficulty of obtaining an unduplicated count of eligible children in Head Start, various child care programs, state prekindergarten, and so on. The NHES:93 and NHES:95 (program participation interview) inquire separately about children's enrollment in Head Start and other center/nursery school/preschool/prekindergarten programs. The child module of the SIPP employs the same strategy of distinguishing Head Start from other group programs. Currently, respondents are asked only about the two most frequently used arrangements, so some programs could be missed. Future rounds of data collection for the SIPP will inquire about all arrangements.

The OECD data contain an indicator titled, "net rates of participation in early childhood education". Based on data provided by each participating country, this indicator includes information on the age and number of years in which children aged 2 to 6 years typically participate in early childhood education. If there is interest in assuring that our indicators facilitate international comparisons, this data source is worth exploring.

For the future, the proposed Survey of Program Dynamics is a logical focal point for the collection of relevant indicators. We also suggest that the availability and validity of state- and district-level data on prekindergarten enrollments be explored given that the majority of states now supplement federal programs with state subsidized prekindergarten programs for low-income children.

Proportion of Children Under age 13 in Latchkey Situations. As of 1991, more than 1.6 million 5 to 14 year olds regularly spent time alone before and after school (Casper et al., 1994). The research literature on the developmental effects of self- or latchkey care is not wholly consistent. Children in suburban settings, where self-care is most common, do not appear to be harmed when left to care for themselves after school, although longitudinal evidence for outcomes such as substance abuse and sexual activity is not available. To the extent that negative effects are found, they tend to be restricted to urban samples and children under age 13. Thus, we propose that the proportion of children under age 13 in latchkey situations be included among the list of child care indicators. It is worth noting that we consider this indicator to be closely tied to the issues of parental supervision and involvement that are discussed in the sections on readiness and schooling.

Data are currently available for this indicator. The child care module of the SIPP provides national data on the number of children of employed mothers who cared for themselves for some part of the time their mothers were working. Ideally, we would obtain data regarding "unsupervised time," for all children and for all hours. This hope may materialize when the SIPP child care module expands to include non-working mothers and non-work hours.

Costs of Care Relative to Family Income. As a nation, it is important to consider whether we are tolerating extremely inequitable situations, particularly with respect to parents' ability to secure the resources needed for their children's healthy development. Concern is widespread regarding children's access to health care, for example. We submit that it is also important to consider equity of access to child care, particularly in light of current evidence of wide disparities in child care expenditures between poor and non-poor families. As of 1991, employed mothers living in poverty who paid for child care spent an average 27% of their monthly family income on it, compared with 7% for non-poor women (Casper et al., 1994; Hofferth, Brayfield, Deich, & Holcomb, 1991).

We propose that available data regarding the proportion of families paying for child care and, among those who pay, the proportion of family income spent on child care be included among the indicators of child care that we track at the federal level. These data are currently available from the child care module of the SIPP. The NHES:95 also includes questions about child care fees.

Parental Choice. Recent evidence suggests that the success with which welfare-dependent mothers complete job training and placement programs hinges in part on their perceptions that their children are in child care arrangements of their choice. Meyers and her colleagues report that mothers in California's GAIN program who wished they could use a different child care provider were over twice as likely to drop out of the program than were mothers who were satisfied with their provider (Meyers, 1993). Further, substantial evidence has now documented that about one-quarter of all mothers using child care wish they could change arrangements and that up to half of low-income, single or teen-age mothers report a desire to change.

These data not only document surprisingly high levels of dissatisfaction with child care, and reveal another possible source of income-based inequity in our child care system, but they point to a useful approach for assessing satisfaction with care--an issue that has eluded effective assessment for years. When asked directly about satisfaction with child care, the vast majority of mothers report that they are highly to quite satisfied. These recent data suggest that we may obtain more valid reports from mothers if we ask them whether they are using arrangements that constitute their first choice, or whether they would prefer to change providers.

We are not aware of any nationally representative data sources that assess parents' reliance on child care of their choice, and with which they are comfortable. Plans are in place, however, to add such a question to the child well-being module of the SIPP. The same could be done with the NHES 1996 parent survey.

Priority Indicators

We propose three indicators for our short list. First, we include the proportion of eligible children who receive early intervention programs. Our rationale is threefold: (1) substantial federal and state resources are spent on early intervention programs, (2) these programs appear to offer higher quality care than the typical child care arrangements that low-income children receive, and (3) they can reap positive outcomes for children. We would not want this indicator to provide an incentive to "water down" the quality of these services in order to serve yet more eligible children.

To guard against this, we include indicators of quality of care on our priority list despite the lack of an available data source. This should be a top priority for "future prospects" with serious attention paid to both a downward extension of the ECLS to provide national data and an exploration of state-level child care licensing data to provide an initial state and sub-state indicator. Until reliable indicators of child care quality become available--a long range goal--we propose that improved questions aimed at capturing the stability and choice of care be added to the SIPP and NHES. These data could serve as interim indicators that are closely linked to children's well-being in child care.

Third, to assure attention to issues of equity, we include the proportion of family income spent on child care as a family-level indicator that is currently available.

EARLY SCHOOLING

There is no agreed upon demarcation between the assessment of school readiness and that of schooling outcomes, although it is logical to consider the post-kindergarten years as falling within the purview of schooling indicators. Here again, other papers in this volume are highly relevant, most notably that on achievement outcomes, but also those by Love, Aber, and Brooks-Gunn that capture non-academic aspects of schooling.

We focus our discussion on the important transition that characterizes schooling around grades three and four--often considered the transition from primary to elementary school--for two reasons. First, the school curriculum undergoes an important shift at this stage from one that emphasizes the acquisition of skills (reading, writing, computation) to one that begins to emphasize the use of these skills (e.g., reading for comprehension, writing for communication functional uses of numbers). Accordingly, it is our recommendation that indicators of schooling, focused on this period, attempt to capture the functional uses of knowledge, rather than just the amount of knowledge that a given child has acquired.

Second, this half-way point between school entry and middle school has been identified as a particularly vulnerable period for schooling outcomes. Labeled the "third grade slump," it is not uncommon for some children who have been performing adequately through second grade to experience decrements in achievement around third and fourth grade, presumably as a result of the change in pedagogy.

Indicators for Early Schooling

Once children enter the elementary school years, aggregate data are available to assess achievement and school functioning. Although we focus on third-fourth graders, we emphasize topics that we believe have relevance for all school-age children. Of course, the specific indicators that capture each topic will vary with the age of the child (e.g., a child's engagement in school will manifest itself somewhat differently in elementary school and in high school). There is also fairly wide agreement about aspects of schooling that are important to capture. These include: (1) Achievement, (2) Progress in school (proportion of children at grade level, rates of grade failure/retention, placements in remedial classes or gifted classes, receipt of special education services), (3) Engagement in school (absenteeism, extracurricular activities), and (4) Parental involvement/participation. We also discuss bilingualism as a potentially important area for the future development of indicators. Although we focus on national data, schooling indicators are particularly amenable to documentation with state and local data--an important issue for future exploration. Table 3 presents our proposed list of indicators of early schooling.

Achievement. Indicators of school achievement should be linked to the educational goals of the nation, and whatever assessments are used to track progress on these goals should be among our national indicators. As these assessments are being developed, the National Assessment of Education Progress (NAEP) offers the most obvious source of indicator data regarding student achievement, particularly given its fourth grade starting point.

We appreciate the difficult debate that has recently accompanied efforts to set achievement levels in conjunction with the NAEP (e.g., basic, proficient, advanced levels of achievement), but consider it important to work towards some indicator that captures criterion-referenced levels of knowledge. We further encourage consideration of functional measures of achievement, beginning at the third-fourth grade level and continuing through-out the child's schooling. By "functional" we mean to capture the difference between having acquired a body of knowledge (e.g., knowing how to read and comprehend text) and putting this knowledge to use (e.g., using books to acquire knowledge and understanding).

Progress in school. Beyond measures of achievement, school functioning is most commonly assessed through measures of grade failure/retention, placements in remedial classes or gifted classes, and receipt of special education services. These measures track the share of children who are showing patterns of progress in school that depart from the typical range. They are another obvious candidate for inclusion in a set of schooling indicators. One of the major decisions to be made concerns the source of data on which this project should rely: parents, teachers, records. Ideally, convergent evidence from multiple data sources would be used.

Among the current sources of data are: (1) the NHES:93 asks parents about grade repetition and receipt of special help (we are unsure about whether gifted classes are included), (2) the Child Module of the SIPP asks parents about grade repetition, placement in gifted classes, and school suspension, (3) the Profiles study also contains relevant items on children in schools in low-income districts. The Schools and Staffing Surveys may also contain pertinent information.

In the future, the ECLS will surely include information on children's progress in school. In addition, the NHES:95 includes comparable information to that in the 1993 protocol, and the proposed Survey of Program Dynamics will collect relevant data.

Engagement in school. This construct becomes much more important once the child passes beyond the elementary years, and can be assessed with negative measures of absenteeism, as well as with positive measures of participation in extracurricular activities and special roles in school. At the elementary level, absenteeism is important to track because it has a direct effect on children's opportunity to learn. Variation in attendance, however, is probably affected more by health and other factors beyond the child's control than by the child's interest in school during these early years.

The ECLS will likely be a useful source of data on young children's school attendance. For current data, the NHES:93 School Safety and Discipline component (interviews with parents of children in grades 3 through 12 and youth in grades 6 through 12) includes information about participation in extracurricular activities, suspensions, and problems in school, as does the SIPP Child Well-being Module. The proposed Survey of Program Dynamics would also collect such information.

Parental Involvement/Participation. Parental involvement in schooling is actually a somewhat ill-defined construct. Although it is difficult to be "against" parental involvement, the literature in this area remains unclear about exactly what forms and amounts of parental involvement really matter for children. To illustrate the conceptual confusion, consider the time parents spend helping the child with homework. It is entirely possible that parents who spend relatively high amounts of time involved with homework will have children who do relatively well in school. Alternatively, at least some parents who provide substantial help with homework may do so because their child is doing poorly in school or is resisting homework. Another issue that generates debate is whether it is involvement with the child at home or involvement in the school setting (or both, since they are likely correlated) that constitutes the most predictive form of involvement. Finally, there may be cultural differences in how parents express their commitment and engagement with their child's schooling.

Nevertheless, for the reasons discussed above (see readiness), we propose that patterns of parent involvement in schooling are important to capture. Recent evidence that during the post-elementary years parent involvement in the child's school setting predicts children's academic standing, classroom conduct, and rates of suspension even when related family factors are controlled (Zill & Nord, 1994), provides additional support for our position. Relevant behaviors would include the parents' familiarity with the child's teacher, their perceptions of the school's receptivity to their involvement, number of times they have visited the child's teacher/classroom (school and non-school hours) for positive or routine reasons (excluding visits occasioned by the child's negative conduct), attendance at PTA and other policy oriented meetings, and other roles assumed in conjunction with the school (e.g., volunteer work, parent committees). Unfortunately, we are not aware of any current data sources that inquire about relevant behaviors, although the NHES 1996 parent interview may include relevant information. The ECLS may provide a vehicle for the development of indicators of parental involvement and the Survey of Program Dynamics may offer a future source of on-going indicator data.

Bilingualism. As discussed above (see child care indicators), today's children will need to be prepared to achieve, contribute, work, and parent in a multicultural society and global economy. Exposure to a language other than English (for English speakers) and support for native languages among children whose first language is not English strikes us as very basic indicators of the extent to which our nation's schools are preparing children for this future. Thus, we recommend inclusion of an indicator of children's exposure to non-English instruction among the set of schooling indicators that are developed in conjunction with this new initiative. The NHES:95 may be a source of pertinent information and the OECD indicators should also be reviewed with this indicator in mind.

Priority Indicators

We propose that three indicators be included on the priority list: (1) Achievement, (2) Progress in school, and (3) Parent involvement. The first two are probably non-controversial; we include the third because it focuses on a positive outcome, embraces a family-level indicator, and is receiving growing empirical support as an important predictor of children's school achievement.

Conclusions

This paper covers a wide territory. We have focused more on the early childhood and "readiness" stages of development than on the indicators of schooling during the elementary years. We strong suggest that our recommendations for indicators be placed in the context of other papers in this volume that focus on schooling outcomes and encompass the full range of factors--cognitive, social, health--that predict and reflect success in school.

As a final note, we address the question of a schedule for indicators data collection. Our approach emphasizes major developmental and institutional transitions in children's lives. These occur when children first encounter school-like settings and curricula, first enter formal school settings, and at the third-fourth grade (see discussion above). Therefore, we recommend collection of readiness, child care, and early schooling data at these transition points, for which it is important to recognize that chronological age is an imperfect proxy. This translates into assessments of 3-year olds (when a substantial share of children in out-of-home settings are in group care/education settings), kindergarten-age children (at kindergarten entry), and third- or fourth-graders.

References

Aber, J.L, Molnar, J., & Phillips, D. (1986). Action research in early education. Report prepared for the Foundation for Child Development. New York: Foundation for Child Development.

Alexander, K.L., & Entwisle, D.R. (1988). Achievement in the first 2 years of school: Patterns and processes. Monographs of the Society for Research in Child Development, 53 (No. 2, serial no. 218). Chicago: University of Chicago Press.

Benasich, A.A., Brooks-Gunn, J., & McCormick, M.D. (1992). Behavioral problems in the two-to-five year old: Measurement and prognostic ability. Unpublished manuscript. Princeton, NJ: Educational Testing Service.

Boyer, E. (1991). Ready to learn: A mandate for the nation. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

Bradburn, N.M. (July, 1994). Outline of issues report for Early Childhood Educational Study. (Draft copy). Chicago, IL: National Opinion Research Center.

Bradley, R.H., Caldwell, B.M., & Rock, S.L. (1988). Home environment and school performance: A ten-year follow-up and examination of three models of environmental action. Child Development, 59, 852-867.

Brandon, R.N. (March, 1992). Social accountability: A conceptual framework for measuring child/family well-being. Paper prepared for the Center for the Study of Social Policy. Washington, D.C.: CSSP.

Brayfield, A.A., Deich, S.G., & Hofferth, S.L. (1993). Caring for children in low-income families. A substudy of the National Child Care Survey, 1990. Washington, D.C.: Urban Institute.

Bronson, M.B. (1994). The usefulness of an observational measure of young children's social and mastery behaviors in early childhood classrooms. Early Childhood Research Quarterly, 9, 19-43.

Case, R., & Griffin, S. (1990). Child cognitive development: The role of central conceptual structures in the development of scientific and social thought. In C-A. Hauert (ed.), Developmental Psychology: Cognitive, Perceptuo-Motor and Psychological Perspectives. Amsterdam, The Netherlands: Elsevier.

Casper, L., Hawkins, M., & O'Connell, M. (April, 1994). Who's minding the kids? Statistical Brief (SB/94-5). Washington, D.C.: U.S. Department of Commerce.

Cazden, C.B., Snow, C.E., & Heise-Baigorria, C. (July, 1990). Language planning in preschool education. Unpublished document prepared for the Consultative Group on Early Childhood Care and Development. New York: UNICEF.

Chang, H. (1993). Affirming children's roots: cultural and linguistic diversity in early care and education. San Francisco, CA: California Tomorrow.

Coleman, M., Robinson, B., & Rowland, B. (1990). Children in self-care: An examination of research confounds. Children and Youth Services Review, 12, 327-339.

Crouter, A.C., MacDermid, S.M., McHale, S.M., & Perry-Jenkins, M. (1990). Parental monitoring and perceptions of children's school performance and conduct in dual- and single-earner families. Developmental Psychology, 26, 649-657.

Cummings, M.E. (1980). Caregiver stability and day care. Developmental Psychology, 16, 31-37.

Dickinson, D. K., and D. E. Beals (1994). Not by print alone: Oral language supports for early literacy development. In DeLancy (Ed.), *Emergent literacy: From research to practice*. New York: Praeger Publishers.

Dishion, T.J. (1990). The family ecology of boy's peer relations in middle childhood. *Child Development*, 874-892.

Entwisle, D.R., & Alexander, K.L. (1990). Beginning school math competence: Minority and majority comparisons. *Child Development*, *61*, 454-471.

Entwisle, D.R., Alexander, K., Pallas, A., & Cadigan, D. (1987). The emergent academic self-image of first graders. *Child Development*, *58*, 1190-1206.

Galinksy, E., Howes, C., Kontos, S., & Shinn, M. (1994). *The study of children in family child care and relative care*. New York: Families and Work Institute.

Goldenberg, C. (1987). Low-income Hispanic parents' contributions to their first-grade children's word-recognition skills. *Anthropology and Education Quarterly*, *18*, 149-179.

Griffin, S., Case, R., & Sandieson, R. (1992). Synchrony and Asynchrony in the acquisition of everyday mathematical knowledge: Towards a representational theory of children's intellectual growth. In R. Case, (ed.), *The Find's Staircase: Exploring the Central Conceptual Underpinnings of Children's Theory and Knowledge*. Hillsdale, NJ: Erlbaum Press.

Griffin, S., Case, R., & Siegler, R. (1992). *Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure*. Unpublished manuscript, Department of Education, Clark University.

Hakuta, K., & D'Andrea, D. (1992). Some properties of bilingual maintenance and loss in Mexican background high-school students. *Applied Linguistics*, 13:72-99.

Hayes, C., Palmer, J., & Zaslow, M. (Eds.) (1990). *Who cares for America's children: child care policy for the 1990s*. Washington, D.C.: National Academy Press.

Helburn, S., Culkin, M.L., Morris, J., Mocan, N., Howes, C., Phillipsen, L., Bryant, D., Clifford, R., Cryer, D., Peisner-Feinberg, E., Burchinal, M., Kagan, S.L., & Rustici, J. (January 1995). *Cost, Quality, and Child Outcomes in Child Care Centers*. Executive Summary. Boulder, Colorado: University of Colorado.

Hofferth, S.L., Brayfield, A., Deich, S., & Holcomb, P. (1991). *National child care survey, 1990*. Washington, D.C. : Urban Institute Press.

Howes, C. (1988). Relations between early child care and schooling. *Developmental Psychology*, *24*, 53-57.

Kagan, S.L. (December, 1990). Readiness 2000: Rethinking rhetoric and responsibility. *Phi Delta Kappan*, 272-279.

Kagan, S. L., and Garcia, E. (1991). Educating culturally and linguistically diverse preschoolers: Moving the agenda. *Early Childhood Research Quarterly*, 6:427-443.

Gresham, F., & Elliott, S. ('990). *The social skills rating system*. MN: American Guidance Service.

Layzer, J.I., Goodson, B.D., & Moss, M. (1993). *Life in preschool. Final report of the Observational Study of Early Childhood Programs (Volume 1)*. Washington, D.C.: U.S. Department of Education.

Lee, V., Brooks-Gunn, J., Schnur, E., & Liaw, T. (1990). Are Head Start effects sustained? A longitudinal comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. Child Development, 61, 495-507.

Long, L. & Long, T. (1981). Latchkey children: The child's view of self care. Washington, D.C.: Catholic University. ERIC Document Reproduction Service, ED 211 229.

Love, J.M., Aber, J.L., & Brooks-Gunn, J. (October, 1994). Strategies for assessing community progress toward achieving the first national educational goal. Princeton, NJ: Mathematica Policy Research, Inc.

Meisels, S.J., Wiske, M.S., Henderson, >IW., Marsden, D.B., & Browining, K.G. (1988). ESI 4-6: Early Screening Inventory for Four Through Six Year Olds. Revised, third edition. New York: Teachers College Press.

Meyers, M.K. (1993). Child care in JOBS employment and training program: What difference does quality make? Journal of Marriage and the Family, 55.

Mischel, W. (1984). Convergences and challenges in the search for consistency. American Psychologist, 39, 351-364.

Moore, K. (1994). New social indicators of child and family wellbeing for the twenty-first century. (draft paper). Washington, D.C.: Child Trends, Inc.

Moorehouse, M.J. (1991). Linking maternal employment patterns to mother child activities and children's school competence. Developmental Psychology, 27, 295-303.

National Center for Education Statistics (September 1991). Education Counts: An Indicator System to Monitor the Nation's Educational Health. Report of the Special Study Panel on Education Indicators. Washington, D.C.: Department of Education.

National Education Goals Panel (December, 1993). Reconsidering children's early development and learning: Toward shared beliefs and vocabulary. Draft report to the National Education Goals Panel from the Goal 1 Technical Planning Group. Washington, D.C.: NEGP.

NICHD Early Child Care Network (1994). Child care and child development: The NICHD Study of Early Child Care. In S.L. Friedman & H.C. Haywood (eds.), Developmental Follow-up: Concepts, domains, and methods. (pp. 373-396). San Diego, CA: Academic Press, Inc.

Office of Educational Research and Improvement (1994). Reinventing the office of educational research and improvement. Washington, D.C.: OERI.

Organization for Economic Co-operation and Development (1993). Education at a glance: OECD indicators. Paris, France: OECD.

Phillips, D., & Crowell, C. (1994). Cultural diversity and early education. Washington, D.C.: National Academy Press.

Phillips, D., Voran, M., Kisker, E., Howes, C., & Whitebook, M. (1994). Child care for children in poverty: Opportunity or inequity? Child Development, 65, 472-492.

Schorr, L.B. (1994). The case for shifting to results-based accountability. Paper written for the Improved Outcomes for Children Project. Washington, D.C.: Aspen Institute.

Siegler, R.S., & Robinson, M. (1982). The development of numerical understandings. In H.W. Reese and L.P. Lipsitt (eds.), Advances in Child Development and Behavior. NY: Academic Press.

Snow, C. E. (1992). Perspectives on second-language development: Implications for bilingual education. Educational Researcher, 21:6-19.

Snow, C.E. (1983). Literacy and language. Harvard Educational Review 53, 165-189.

Stanford Working Group (1993). Federal education programs for Limited-English-Proficient students: A blueprint for the second generation. Stanford, CA: Stanford University.

Stipek, D. (1988). Motivation to learn: Theory to practice. Englewood Cliffs, NJ: Prentice-Hall.

U.S. Department of Education (1992). The Condition of Bilingual Education in the Nation: A Report to the Congress and the President, June 30, 1992.

Vandell, D., & Posner, J.K. (in press). Low-income children's after school care: Are there beneficial effects of after school programs? Child Development.

Walker, H.M., Block-Pedego, A., Todis, B., & Stevenson, H. (1991). School Archival Records Search (SARS). Longmont, CO: Sopris West Inc.

West, J. (1994). Early Childhood Longitudinal Study Kindergarten Cohort. Washington, D.C.: National Center for Education Statistics.

Whitebook, M., Howes, C. & Phillips, D. (1989). Who cares? Child care teachers and the quality of care in America. Final report of the National Child Care Staffing Study. Oakland, CA: Child Care Employee Project.

Zill, N., & Nord, C.W. (1994). Running in place: How American families are faring in a changing economy and an individualistic society. Washington, D.C.: Child Trends, Inc.

Zill, N., Stief, T., & Coiro, M.J. (1992). Scalability and reliability of home environments items from the 1991 National Household Education Survey. Unpublished manuscript. Washington, D.C.: Child Trends, Inc.

Table 1

Indicators of School Readiness

Indicator	Current Sources	Future Prospects
Exposure to Reading at Home	NHES:93	NHES:95/96 SIPP Child Module NLSY-MC ECLS
Exposure to Pre-Numeracy Experiences		NHES:95/96 ECLS
Approaches to Learning		ECLS Love et al., 1994
Emergent Literacy and Numeracy Development	NHES:93 Prospects Study	NHES:95/96 SIPP Child Module ECLS Love et al., 1994 State/local level data
Proportion of Kindergartners "Unready" for Kindergarten	NHES:93	NHES:95/96 SIPP Child Module State/local level data
Parental Attitudes/Expectations		NHES:95/96 ECLS Survey of Program Dynamics
Access to Instruction in Native Language		NHES:95 OECD Schools and Staffing Study

Table 2

Child Care Indicators

Indicator	Current Sources	Future Prospects
Quality of Care		ECLS State regulatory data
Stability of Care	SIPP Child Care Module	NHES:95 SIPP Child Module
Proportion of Eligible Children in Early Intervention Programs		Survey of Program Dynamics State/local level data
Proportion of children in Latchkey Situations	SIPP Child Care Module	
Child Care Costs: Family Income	SIPP Child Care Module	NHES:95
Parent Choice	SIPP Child Module	NHES:95/96

Table 3

Indicators of Early Schooling

Indicator	Current Sources	Future Prospects
Achievement	NAEP	ECLS NEGP initiatives State/local level data
Progress in School	NHES:93 Profiles Study	SIPP Child Module NHES:95/96 ECLS Survey of Program Dynamics
Engagement in School	NHES:93 SIPP Child Module	ECLS Survey of Program Dynamics
Parental Involvement/ Participation		NHES:96 ECLS Survey of Program Dynamics
Bilingualism		NHES:95 OECD ECLS

Indicators of High School Dropout

Robert M. Hauser
Department of Sociology
Institute for Research on Poverty
University of Wisconsin-Madison

November 1994

Prepared for the Conference, Indicators of Children's Well-Being, Rockville, Maryland. This research was supported in part by grants from the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services; from the National Institute on Aging; and from the Spencer Foundation. It was carried out using facilities of the Center for Demography and Ecology at the University of Wisconsin-Madison, for which core support comes from the National Institute of Child Health and Human Development, and facilities of the Institute for Research on Poverty, which is supported by a grant from the Office of Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. I thank Linda Jordan, James A. Dixon, Taissa S. Hauser, Julia Gray, and Yu Xie for assistance in the preparation and documentation of the Uniform October Current Population Survey file, 1968-1990. Those data are available from the Interuniversity Consortium for Political and Social Research. Please direct all correspondence to Robert M. Hauser, Department of Sociology, The University of Wisconsin-Madison, 1180 Observatory Drive, Madison, Wisconsin 53706 or HAUSER@SSC.WISC.EDU. The opinions expressed herein are those of the author.

Abstract

Spurred by our National Educational Goals, as well as by the economic woes of school dropouts, we have improved the conceptualization and measurement of high school dropout. Moreover, greater public and private resources have been devoted to development and dissemination of temporally and spatially comparable indicators of dropout. These may help improve both our understanding of school-leaving and the allocation of resources to prevent or remediate the effects of dropout. At the same time, indicators of dropout are sometimes weak, misleading, or excessively aggregated. Recent changes in the Census concept of educational attainment have in some ways made it more difficult to identify high school dropouts, and many indicator series fail to identify proximate sources of school-leaving. At the national level, improvement should focus on timely measurements for major social and economic groups that will add theoretical understanding to indicator time series. At the state level, there are painful trade-offs between timeliness, specificity, and validity; model-based estimates may lead to improvement over present practice. Model-based estimates also offer possibilities for creation of risk-adjusted series.

Indicators of High School Dropout

The highly publicized National Goals for Education (U.S. Department of Education 1990) have proclaimed 90 percent high school completion by the year 2000 among six primary goals. It is not clear what "90 percent" means in this context. A recent report of the Department of Education (Tomlinson, Frase, Fork, and Gonzalez 1993:2) notes uncertainty about what marks high school graduation, how the goal can be reconciled with state-to-state variation in graduation requirements, and what populations, at what ages, should be defined as at risk of graduation.[1] These issues, among others, also arise in the measurement of high school dropout, but issues of validity, reliability, and feasibility come into play, as well as those of politics and administration.

The latest report of the National Educational Goals Panel (1994), states the goal is to "increase the percentage of 19- and 20-year-olds who have a high school credential to at least 90 percent," and the report finds that high school completion is one of six areas in which "no significant changes in national performance have occurred." As of 1992, "the nation is already very close to achieving the 90 percent target," for 87 percent of 19- and 20-year-olds have completed high school. Completion rates were 91 percent among whites, 81 percent among Blacks, and only 65 percent among Hispanics, so we must "make serious efforts to close the persistent gap in completion rates between White and minority students." I shall comment below about conceptual and statistical aspects of this target.

Some Consequences of High School Dropout

While the public perception of high school dropout as a social problem has been widespread for at least 30 years (Schrieber 1967), recent years have brought increasing evidence that the failure to complete high school is associated with problems in employment, earnings, family formation and stability, civic participation, and health. For example, Figure 1 shows trends and differentials in employment rates of persons 25 to 34 years old by sex and educational attainment from the early 1970s to the early 1990s.[2] In every year and among women and men, employment varies directly with completed schooling. Moreover, there appears to be a growing differential in employment between dropouts (here defined as those with 9 to 11 years of schooling) and either high school or college graduates. The sources of the growing differential are different among men and women. Among men, employment has been very high and stable among college graduates, while it has declined, both among high school graduates and, to an even greater extent, among dropouts. Among women, employment has increased among dropouts as among all women, but the growth has been much greater among high school and college graduates. In the early 1970s, about 30 points separated the chances that a male dropout and a woman college graduate would be employed. By the early 1990s, a college woman was about 10 percentage points more likely to work outside the home than was a male dropout.

Just as the earning power of high school graduates has declined relative to that of college graduates (Murphy and Welch 1989; Murnane and Levy 1993; Hauser 1993), so has the earning power of high school dropouts relative to high school graduates. Indeed, in many cases, high school dropouts are already unable to compete for jobs that pay enough to keep one out of poverty. The economic consequences of dropping out of high school have never been so severe. Among men and women wage and salary workers, dropouts make substantially less than high school graduates (Figure 2 and Figure 3).[3] Over the past two decades, the earnings of white male dropouts have declined from 85 percent to less than 75 percent of the earnings of white high school graduates.[4] Among African-American and Hispanic men, the time series is far more variable, but there also appears to be some evidence of a decline in earnings relative to high school graduates. Among women, there is no obvious long term trend in the relative earnings of high school dropouts, but the differential fluctuates around a level of 0.6. That is, women high school graduates earn about two thirds more than dropouts.

---

[1] This uncertainty is, perhaps, no worse than that surrounding the level of health care coverage that various politicians were willing to call "universal." Another competitor in this league is the stated goal of reducing differential mortality in OECD countries by 25 percent.

[2] The employment rate is just the ratio of employed persons to the total population in the specified group; that is, it ignores labor force status. These persons are old enough so differentials in age between recent dropouts and graduates should not much affect employment differentials; indeed, for dropouts and graduates of the same age, experience is inverse to schooling.

Across the past three decades, the odds of voting in Presidential elections have increasingly favored those who have graduated from college or at least completed part of a college education relative to high school graduates (Figure 4). At the same time, the chances of electoral participation by high school dropouts have decreased relative to those of high school graduates. Obviously, illustrative differentials between dropouts and graduates could be spun out endlessly. The bottom line is that the failure to obtain at least a high school diploma looks more and more like the contemporary equivalent of functional illiteracy. It suggests a failure to pass minimum thresholds of economic, social, or political motivation, access, and competence.

Trends in High School Dropout

Since the middle 1980s, there has been a steady stream of new reports about the familial and economic origins of high school dropout (McLanahan 1985; Ekstrom, Goertz, Pollack, and Rock 1986; Krein and Beller 1988; Astone and McLanahan 1991; Haveman, Wolfe, and Spaulding 1991; Sandefur, McLanahan, and Wojtkiewicz 1992; Hauser and Phang 1993). The National Center for Education Statistics now produces a regular series of annual reports on trends and differentials in high school dropout (Frase 1989; Kaufman and Frase 1990; Kaufman, McMillen, and Whitener 1991; Kaufman, McMillen, Germino-Hausken, and Bradby 1992; McMillen, Kaufman, Germino-Hausken, and Bradby 1993; McMillen, Kaufman, and Whitener 1994). Thus, the association of high school dropout with educational and economic deprivation, minority status, and family disruption is well documented, as are global trends in various measures of high school dropout.

Overused as it may be, Dickens' wonderful opening line, "It was the best of times, it was the worst of times," neatly encapsulates public views about high school dropout.[5] At the least, the times are confusing. According to the Children's Defense Fund (1994: xii), "Every 5 seconds of the school day a student drops out of public school." Moreover, "no significant progress has been made nationally since 1985 in reducing the proportion of students who drop out before completing high school. In 1991, 12.5 percent of all young people ages 16 through 24 who were not enrolled in school did not have a high school diploma or its equivalent, up slightly from 12.1 percent in 1990" (p. 55).[6] This text garbles the concept: It should read "In 1991, 12.5 percent of all young people ages 16 to 24 were not enrolled in school and did not have a high school diploma or its equivalent." As a matter of possible interest, I have recreated this series, by age, from 1978 to 1993 in Figure 5. Within age subgroups, there would appear to be some year-to-year unreliability, but it would be hard to ignore the overall downward trend; the rise in dropout noted by the Children's Defense Fund from 1990 to 1991 was more than reversed between 1991 and 1992.[7]

---

[3] Graduates here are individuals with exactly 12 years of schooling or a high school diploma or equivalent.

[4] The earnings of white male high school graduates have also declined in real terms.

[5] The full passage is *a propos*: " ... it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way" (Charles Dickens, *A Tale of Two Cities*, Book 1, Chapter 1).

[6] This text garbles the concept: It should read "In 1991, 12.5 percent of all young people ages 16 to 24 were not enrolled in school and did not have a high school diploma or its equivalent." As a matter of possible interest, I have recreated this series, by age, from 1978 to 1992 in Figure 1.

One might draw the same conclusion as the Children's Defense Fund by consulting another source, the annual *Kids Count Data Book* of the Annie E. Casey Foundation (1994). One of ten key *Kids Count* indicators, percent graduating from high school on time, declined regularly from 71.6 percent in 1985 to 68.8 percent in 1991.[8] In its state rankings, higher on-time graduation rates are presumably better than lower on-time graduation rates, yet the overall rate of on-time graduation declined nationally, even as some other indicators of high school completion were increasing.

The latest edition of the *Condition of Education* (National Center for Education Statistics 1994) opens with a classic, "good news, bad news" story. It offers as good news that, "Overall high school dropout rates have gradually decreased. The differences between dropout rates for blacks and whites have also narrowed. ... This is encouraging because schools provide young people with the opportunity to explore their interests and develop their talents. It is also encouraging because staying in school is an important indication that a young person is learning to be a productive member of U.S. society and is less likely to suffer from poverty and unemployment" (p. iii).

Indeed, the percentages of high school students in grades 10 to 12 who persisted, either by enrolling in two successive years or completing high school, grew steadily from a trough of 93.3 percent in 1978 up to 95.5 percent in 1993.[9] The annual persistence rates began to increase in the late 1970s, and the rates among young men and women have followed very similar trajectories. If anything, women appear to have enjoyed greater chances of remaining in high school than men through most of the past 20 years (Figure 6). Among whites, the annual persistence rate grew from 94.2 percent to 96.1 percent, and among African-Americans, it grew from 89.8 percent to 94.2 percent (Figure 7). Only among Hispanics does there appear to have been little or no sustained growth in this measure of school retention. Moreover, the gap between youth from high and low income families declined from the late 1970s to the middle 1990s. In 1978, year-to-year persistence was 82.9 percent in the lowest fifth of family income and 97.0 percent in the highest fifth of family income (Figure 3). By 1993, persistence was 87.7 percent in the lowest fifth of family income and 98.7 percent in the highest fifth of family income (National Center for Education Statistics 1994: 32; McMillen, Kaufman, and Whitener 1994: 6).

Are high school dropout rates really rising or falling? Or is there no clear pattern? To make matters short, my reading of the evidence is more consistent with that of the NCES than that of the Children's Defense Fund, but it is worth thinking carefully about the conflicting interpretations one might draw from alternative measures of school leaving. In this discussion, I examine and compare some of the measures of high school completion or non-completion that are now in wide use and offer my suggestions for improving them. I shall not attempt to lay out a detailed set of criteria for good indicators of high school dropout, but some of the important desiderata are as follows. First, an indicator should have face validity. A positive indicator should rise when conditions are improving and fall when they are getting worse. Second, an indicator should be conceptually sound. It should be consistent with a reasonable understanding of the process or processes that it purports to measure. It should pertain to a well-defined set of persons or events. It should be understandable to the public at large. Third, it must be possible to ascertain the indicator comparably for different populations, e.g., for different populations, political or administrative units, or time periods. Fourth, an indicator should be timely, and in two different ways. It should tell us about outcomes of the educational process as early in the course of schooling as it is feasible to think of them as highly determined, and it should be measured and disseminated soon afterward. Fifth, an indicator should be statistically reliable, so we can know whether things are really getting better or worse. Sixth, it should be possible to analyze the sources of trends and differentials in the indicator. That is, it should not merely be of self-evident diagnostic value, but it should be possible to link the indicator to other relevant data.

---

[7] The recency of the data in the Spring 1994 report of the Children's Defense Fund would appear to be about a year off, relative to the availability of data. The CDF series ends in 1991 yet, by the Fall of 1994, NCES reports dropout rates for 1993 from the same source, the October Current Population Survey.

[8] The *Kids Count Data Book* notes, "This measure is not the same as a dropout rate. Some of those who fail to graduate on time are dropouts, but others are simply falling behind their peers. It is worth noting, however, that those who fall behind age/grade norms are more susceptible to dropping out eventually" (1994: 16).

[9] This statistic is the inverse of an annual dropout rate proposed by Robert Kominski (1990) of the U.S. Bureau of the Census.

Figure 9 shows the trend in high school completion by race-ethnicity and sex over the past three decades. The measure is the percentage of persons aged 25 to 29 who completed 12 or more years of school, as reported in March Current Population Surveys. Presumably, 25 to 29 is old enough to cover almost all persons who complete high school. There are few differences in high school completion by sex within the three largest race-ethnic groups (white, black, and Hispanic). Blacks began this period, in the middle of the civil rights revolution, well below the level of high school completion that Hispanics had achieved in the early 1970s, when we first began to measure their attainments. By the middle 1970s, white attainment stabilized at about 85 percent, and continued growth among African Americans has brought them to nearly the same completion rate as whites.

What is right and wrong with this indicator? It is conceptually sound, one of the measures that the NCES refers to as a "status" indicator. It depends on two counts, the total population of interest (by age) and the total population who completed high school. Thus, the numerator depends on the measurement of educational attainment or years of completed schooling. As measured since the early 1940s, that variable can be measured with high reliability in social surveys, either directly or by proxy (Bielby, Hauser, and Featherman 1977). It is measured regularly, as a control card item, in the monthly Current Population Survey. The wobbly trend lines for blacks and Hispanics in Figure 9, which I have not attempted to smooth out, suggest that, for minority populations, there is substantial statistical unreliability in the CPS measure of attainment. While it is usually reported only for persons in the March Annual Demographic Survey, we could readily increase reliability, either by smoothing the series across years or by pooling data from other months of the CPS, e.g., from the October samples, whose membership never overlaps with that of the March samples. One of the persistent problems in the tabulation and reporting of data from Current Population Surveys is that they are almost always reported in CPS month by calendar year "chunks," and this fails to take advantage of the essential comparability of the surveys from month to month and year to year. Moreover, because educational attainment is, in principle, cumulative and irreversible, we can improve the reliability of historic series by combining data for older and younger persons, classified by age within survey cross-sections. There are two limitations to the last of these possibilities, first, that reports from older individuals tell us about the increasingly distant past and, second, that there appears some tendency for older adults to exaggerate their levels of completed schooling. Thus, efforts to estimate educational differentials by inter-censal survivorship have sometimes observed negative mortality at higher levels of schooling.

One serious problem with this indicator is that it is not timely, at least in the first sense noted above. By the time the population reaches ages 25 to 29, most people are 7 to 12 years beyond the modal age at high school completion. Thus, the measure is, at best, about a decade behind the realities of school progression and dropout. It is valuable when we choose to look back at the progress, or lack thereof, among major social and economic groups. On the other hand, once observed, we can obtain this measure in timely fashion. Data from the March Current Population Survey are regularly available in the fall of the year in which they were collected. Thus, in principle, a decade-old look at dropout trends can be had within six months of its observation.

Another serious issue in the use of this indicator pertains to recent changes in the conceptualization and measurement of educational attainment, both in the U.S. Census of Population and in the Current Population Survey. In the U.S., if not elsewhere, it used to be easy to ascertain educational attainment. It was sufficient to ask, "What was the highest grade of school that ... completed?" and provide numeric categories ranging from zero to 17 or more. In the Current Population Survey (CPS), a most useful distinction was added (Kominski and Adams 1994): "What is the highest grade or year of regular school ... has ever attended? Did ... complete that grade (year)?" This two-part question made it clearer that the question was about regular (academic) schooling. Moreover, it was possible in principle to measure school dropout among those who had attended, but not completed a grade. Over the years, the upper range of responses was expanded, and by 1991, the CPS recorded as many as 26 years of schooling.

Following a similar change in the 1990 Census, the CPS introduced a new, single educational question early in 1992: "What is the highest level of school ... has completed or the highest degree ... has received? The 16 CPS codes and response categories for the new item are displayed in Figure 10 (Kominski and Adams 1994: XIII).[10] The new CPS educational attainment question and its responses differ in several ways from the old item and its response categories. First, it eliminates the probe distinguishing between the highest grade attended and the completion of that grade. Second, responses below the level of secondary school have been grouped. Third, a new category, "12th grade, No Diploma" has been added. Fourth, the category for completion of high school now specifically identifies both high school graduation and obtaining a high school equivalent, such as the GED. Fifth, major changes have been introduced in the classification of schooling beyond high school completion. These are now based on credentials, rather than on the completion of numbers of years of schooling. Among those with less than a college degree, the new system makes three distinctions: Some college but no degree, Associate degree in a technical/vocational program, and Associate degree in an academic program. Of these, the former is described by the Census Bureau as "a residual category." Finally, the new system distinguishes among holders of Bachelor's, Master's, Professional, and Doctoral degrees.

The Bureau of the Census offers four explanations for its construction of the new educational attainment item (Kominski and Adams 1994:XIII-XIV). First, partly because of the increased time from college entry to completion, the presumed equivalency between 16 years of schooling and holding a bachelor's degree has become invalid. The Bureau reports that in the middle 1980s, the traditional equivalency between 16 years of schooling and a Bachelor's degree would over-estimate the number of degree-holders by more than 1 million persons. Second, no specific degrees were identified in the old item. This was a problem, not only because 16 years of school was not equivalent to a bachelor's degree, but because of the increasing importance and prevalence of Associate degrees and post-college degrees. Third, the old system led to uncertainty in the classification of high school graduates because persons who had equivalent credentials were supposed to be counted as completing 12 years of school, but often were not so classified. This has become an increasingly important matter because some localities require a final graduation test or certification; thus, it is possible to complete 12 years of regular schooling without earning a diploma. Fourth, the Census Bureau reports that the old items did not meet specific programmatic needs of federal agencies, especially to ascertain degrees. Because of "a serious space constraint in the decennial instrument," and because "detailed attainment information was not legislatively required (or generally needed) below the fifth grade level," an interagency working group advised the Bureau to collapse categories from the 1st to 4th and 5th to 8th grades (Kominski and Adams 1994: XIV). The latter category was subsequently disaggregated to 5th-6th and 7th-8th at the suggestion of the Bureau of Labor Statistics when the new item was added to the Current Population Survey.

In its discussion of comparability and uses of the old and new CPS education items, the Bureau of the Census discusses three issues (Kominski and Adams 1994: XV). First, there obviously is a break in continuity with the past -- a break in previous time series -- and this is an unavoidable consequence of change. Second, it is no longer possible to use years of schooling as a continuous variable in regression analyses. This is probably just as well, since the effects of schooling (as measured the old way) are often nonlinear, and many analyses are carried out using categorical representations of schooling. The Bureau's discussion does not mention the use of schooling as a dependent variable, where the analysis of ordered categorical variables is more complex. Finally, the Bureau notes that it is necessary to abandon some older summary measures of the level of schooling, such as the median or mean, which have no meaning within the new system. This is no great loss, for such measures were already rendered uninformative by the shape of the educational distribution, which is heavily clustered at or near the completion of high school.

---

[10] I have deliberately included the new numeric codes as well as the descriptions of each category to take note of the Bureau's reason for introducing the new codes, namely, to prevent interviewers from erroneous entries using the previous system of recording highest grade or year.

One might expect that, over time, the new CPS educational attainment item will supplant older items in other surveys, including, but not limited to those carried out by the Bureau of the Census, such as the major health interview surveys. I hope that this will not happen. In my opinion, it was clearly desirable to change to a system in which post-secondary credentials were measured explicitly, but the new CPS item is poorly constructed and fails to obtain important information, including some that was obtained by the old item.

> (1) The collapse of several grade levels below high school has made it impossible to follow age-grade progression at younger ages.[11] Neither can we examine school completion closely among recent immigrant populations or among populations with learning disabilities. In populations with low levels of schooling, the collapse will remain problematic, for example, at older ages or when it is necessary to ascertain educational attainments in past generations.[12]
>
> (2) The failure to distinguish grades attended from grades completed has eliminated our ability to examine a key educational transition, namely, that between college entry and completion of the first year of college. Formerly, persons who had attended, but not completed grade 13 could be classified as early college dropouts. A large number of persons who have ever entered college fall into this category, yet it cannot be measured using the new CPS question. A cross-tabulation of educational attainment under the new and old systems, which was carried out as part of the February 1990 Current Population Survey, shows that 7 million people, more than 20 percent of those classified as having completed "some college, but no degree" under the new system, reported having completed no more than 12 years of school under the old system. Those persons comprise nearly 10 percent of all persons classified as high school graduates by the old item (Kominski and Adams 1994: XVI). This suggests that many individuals who dropped out of college during their first year are now classified as having "completed some college." In other words, the failure of the new item to distinguish sharply between attending and completing a grade has serious consequences.
>
> (3) Similar observations apply to entry and completion of the 12th grade, though I know of no major uses of the distinction between grade attendance and completion in high school. One might suspect that elimination of the "completion" probe accounts in part for the nearly 4 million individuals who are classified by the new item as nominally having "completed" twelve years of school without a high school diploma or its equivalent.[13] This is most problematic, for we do not know whether this category is an artifact of survey methodology or an indication of the application of new standards of academic achievement. Indeed, there are disagreements about how persons with "12th grade no diploma" should be classified. As I understand it, the Census Bureau classifies such persons as non-graduates. In a match between old and new attainment items in the Current Population Survey samples of March 1991 and March 1992,[14] an economic demographer found that 55 percent of persons who reported "12th grade no

---

[11] This problem was exacerbated in the 1990 Census because there was no separate question on the grade level of persons currently enrolled in school. That is, grade level had to be inferred from educational attainment. Because the educational attainment question grouped some levels of schooling and elided the distinction between attending and completing a grade, it was not a suitable tool for the analysis of age-grade progression.

[12] In my study of Wisconsin high school graduates of 1957, the modal schooling level of parents was 8 years. What is it in today's new immigrant populations? I think that it may be particularly important to distinguish between the 7th and 8th grades because the latter denotes completion of elementary school.

[13] A cross-classification of the two items by age might be instructive. That is, if the Bureau's understanding of the sources of the non-certified 12th grade completers is correct, non-certification should occur much more often among younger than older persons.

[14] There is 50 percent overlap, year-to-year, in CPS samples for the same month.

diploma" in 1992 had reported completing 12 years of school in 1991. Thus, Jaeger (1993) recommends combining the new "12th grade no diploma" and "high school graduate" categories into a single category of 12 years of schooling. Taking a statistical compromise, Mare's (1995) analysis of educational trends from 1980 to 1990 allocates the "12th grade no diploma" responses to dropout and completer categories in proportion to their shares in a cross-classification of the two items obtained in the February 1990 Current Population Survey.

(4) The nominal collapse of grades 13 to 15 into "some college no degree" has created a large and extremely heterogenous category, which includes about half as many persons as those classified under either the new or old systems as having completed 12 years of school. The "some college no degree" category contains many more people, 33.2 million, than the 4.3 million in each of the Associate degree categories. That category cannot be compared ordinally to those of persons who completed either a vocational or an academic Associate degree; some completed more and some less post-secondary schooling than the two years usually required for an Associate degree. Not only does the "some college" category contain significant numbers of persons who were classified as obtaining no college education under the old system, but it also contains a modest number of individuals who completed 4 or more years of college education. In fact, each of four grade levels (in the old system) within the "some college no degree" category contains more persons than either of the new Associate degree categories: 12 years (7 million), 13 years (11.5 million), 14 years (9.5 million), and 15 years (4.2 million). This heterogeneity is far greater -- and affects many more persons -- than the heterogeneity in attainment (under the new system) among persons who were classified as having completed 16 years of school under the old system.

(5) Despite its proliferation of categories, the new educational classification fails to distinguish individuals who completed 12 years of school from those who achieved high school equivalency, yet there is strong evidence of differences between regular high school graduates and the growing number of individuals with GEDs (Cameron and Heckman 1992). If the "12th grade no diploma" category makes sense, I think that it makes more sense to place GED holders in that category than to combine them with regular graduates.

(6) Finally, despite, or perhaps because of its failure to measure certification directly, the old educational questions come far closer than the new question to telling us how people spent their time during their formative years. That is, the old educational attainment questions tell us more about the process of growing up than how far a person went in school.

In my judgment, it will be best if the new CPS education questions are not used as a model in other surveys. I hope that the Bureau of the Census will modify its question soon, if possible before rather than in conjunction with the Census of 2000. It is a mystery to me why a novel question that was evidently designed within the severe constraints of the decennial census form need have been adopted with minimal changes in the CPS. I recommend that researchers continue to use the old CPS question to ascertain educational attainment, preferably including the probe about completion of the highest grade attended. I also recommend that a separate question or questions be used to measure the highest diploma, equivalency credential, or degree obtained.[15] Having said all this by way of criticism, I must add that the limited data available thus far show no great aggregate discrepancy in rates of high school completion between the old and new measures.[16]

Might it be possible to ascertain the completion of high school at an earlier stage in the life-history of cohorts and thus overcome the problem of timeliness in attainment measured at ages 25 to 29? Figure 11 provides some evidence about this. I have taken educational attainment from March Current Population Surveys

---

[15] One reasonably good series is used by the National Opinion Research Center in its General Social Survey.

[16] However, the new measure leads to much larger estimates of year-to-year dropout in the 12th grade (McMillen, Kaufman, and Whitener 1994:13).

for six, two-year age groups from 1972 to 1993: ages 19 to 20 (the key age identified by the National Goals Panel), 21 to 22, 23 to 24, 25 to 26, 27 to 28, and 29 to 30.[17] The data are arrayed relative to the year in which each cohort reached ages 19 to 20, so a vertical reading of the overlapping segments of the graph shows the growth of high school completion within a cohort from one age to the next.

I would make three observations about these series. First, two year cohorts are too narrow to provide reliable data from a single Current Population Survey, even for the total population. There is far too much "wobble" in the series displayed in Figure 11. To put this more concretely, the lowest of the three series corresponds to the indicator proposed by the National Goals Panel (but without adjustment for award of high school equivalency). Even for the total population, the standard error of a point estimate of the percentage completing high school is 0.76 percentage points at ages 19 to 20. This is far too large for the measure to be useful in detecting likely true changes from one year to the next. The unreliability is even greater in minority populations: In 1993 the standard errors are 2.59 percentage points among non-Hispanic blacks and 4.31 percentage points among Hispanics (McMillen, Kaufman, and Whitener 1994: 149). Thus, using the rule of thumb that a difference or change of two standard errors is statistically reliable, we should have to observe year-to-year changes of about 2 percentage points in the total population, 7 points among blacks, and 12 points among Hispanics before we could conclude that high school completion had really changed significantly. This indicator is a crude instrument indeed.[18]

In order to render major features of the series more clearly, I have taken three-year moving averages for the period 1963 to 1992. These are displayed in Figure 12, which shows large gaps between levels of attainment at ages 19 to 20 and those at any of the older ages. In my opinion, if a great deal of educational growth occurs, for example, between ages 19 to 20 and 21 to 22, it is questionable whether the younger age is appropriate for the target of our national goal. Moreover, since ages at high school completion are increasing along with attainment levels (Hauser and Phang 1993), it would appear to make sense, for reasons of substance as well as reliability, to extend the upper end of the age range for which high school completion is measured.

There are, also, relatively small gaps between completion levels at ages 21 to 22 and at older ages. Roughly speaking, the gap between ages 19 to 20 and ages 20 to 21 is about as large as that between ages 20 to 21 and ages 29 to 30. That is, even if we must wait until close to age 30 to learn how far a cohort will ultimately go in school, we can learn substantially earlier -- if not by age 20 -- how many in a cohort will complete high school. Thus, following the latest of the NCES reports on high school dropout (McMillen, Kaufman, and Whitener 1994: 51-54), I would suggest that ages 21 to 22 are an appropriate range for assessments of progress in the completion of high school, provided the sample is large enough to yield reliable estimates.

Figure 13 shows another piece of evidence relevant to the age at which high school completion is ascertained. The data are percentages completing high school or more (using the official version of the new Census concept) as ascertained for ages 20 to 24 and 25 to 29 in the Census of 1990 (U.S. Bureau of the Census 1994). Obviously, across states, there is a close relationship between rates of attainment observed at those two ages. The linear correlation is 0.96. Moreover, while a comparison of rates at the two ages confounds maturation with change, it is suggestive that there is very little change in high school completion rates by states between ages 20 to 24 and 25 to 29. The average increase is just 0.4 percentage points. Thus, in order to increase the reliability of subnational comparisons, it would appear reasonable to focus status measures of high school completion on the period when cohorts reach ages 20 to 24.

---

[17] These data are reported by McMillen, Kaufman, and Whitener (1994: Appendix C).

[18] I have based these estimates on the assumption of independence from year to year in samples with equal standard errors. This overstates the level of sampling error, for the overlap of CPS samples from one year to the next reduces sampling variability in estimates of change. On the other hand, even if we were subject to sampling error in only one of two adjacent years, we should have to see a shift of about 1.5 percentage points in the white population, 5.2 points in the black population, and 8.6 points in the Hispanic population before we could conclude that anything had changed from one year to the next.

The *Kids Count Data Book* (1994) measure of timely high school completion would appear to have serious defects, as well as some advantages. The measure is obtained by dividing the number of public high school graduates in the reference year by the public ninth grade enrollment four years earlier, with some adjustments for secondary students not classified by grade and for inter-state migration (p. 160). One advantage is that it is available annually by state. A second advantage is that it is timely, in the sense that it pertains to a standard of accomplishment early in the life of each cohort. On the other hand, there appears to be a substantial lag in the availability of the measure for publication. The 1994 *Kids Count Data Book* reports rates of on time high school completion for 1991.

There are more serious problems with the *Kids Count* measure. I have already noted that it shows steady annual declines during a period when other measures show increasing rates of high school completion (but at older ages). As we approach an upper limit of high school completion, how should we weigh the advantage of timeliness (by age) against eventual completion?

Second, the *Kids Count* measure does not appear to measure the same thing, from state to state, that we observe in percentages of high school completion in the Census. For example, I have looked at state-level correlations between the *Kids Count* rates in 1985 and in 1990 with the Census completion rates at ages 25 to 29 and 20 to 24 (which correspond, roughly, with graduation in 1985 and 1990). These intercorrelations are distressingly low. The four correlations between Census and *Kids Count* measures range from 0.68 to 0.78, which are low indeed at that aggregate level. Moreover, the correlation of the *Kids Count* measures from 1985 to 1990 is 0.89, which suggests either that differential change in aggregate high school completion rates is quite rapid, or that the underlying data may be unreliable. That the latter may be the case is suggested by the inter-annual correlation of 0.94 between the *Kids Count* measures in 1990 and 1991. That is, given an inter-annual correlation of 0.94, if the evolution of the rates were lag-one, the five-year correlation between rates would be $0.94^5 = 0.73$. Since this is far lower than the observed five-year correlation, it would appear likely that the annual observations are unreliable (or that there are year-to-year effects with lags of two or more years).

Third, state-level high school completion rates for all population groups combined do not adequately portray state-to-state differentials in completion among major race-ethnic groups. I have looked at state-to-state correlations among percentages completing high school in the 1990 Census at ages 20 to 24 and 25 to 29 among the five race-ethnic groups recognized in the federal statistical system. The state-level correlations within the same race-ethnic group, but between age groups, are relatively large in the case of the larger race-ethnic groups: 0.96 among non-Hispanic whites, 0.94 among blacks, and 0.93 among Hispanics. They are much lower in the two smaller minority groups: 0.77 among American Indians and 0.65 among Asian and Pacific Islanders.[19] The state level correlations among black and non-Hispanic white rates range from 0.27 to 0.38; the state level correlations among non-Hispanic white and Hispanic rates range from -0.13 to -0.04; and the state level correlations among black and Hispanic rates range from 0.26 to 0.40. Also, the correlations between the rates for non-Hispanic whites and for American Indians range from 0.06 to 0.29, while those between non-Hispanic whites and Asian and Pacific Islanders range from -.05 to 0.14. Given this level of inconsistency in state-level completion rates among the major race-ethnic groups, it is reasonable to wonder how much guidance for public policy is provided by aggregate, annual state-level high school completion data. For example, one might ask whether we can learn more about high school completion in a state among blacks, Hispanics, American Indians, or Asian and Pacific Islanders by looking at current state-level completion rates or by looking at group specific rates in the preceding decennial census.

Annual Dropout or Persistence Rates

The most novel and widely adopted indicator arising from the recent national interest in high school dropout is the annual dropout rate proposed by Kominski (1990, p. 304):

---

[19] I suspect that the low correlations among American Indians and Asian and Pacific Islanders can be attributed to the very small size of these populations in some states.

> By using current and prior enrollment statuses, along with information on years of school completed, it is possible to identify those individuals who were enrolled a year ago, are not enrolled now, and have not completed high school. These individuals are identified as high school dropouts in the past year. The formula for the 1-year rate from grade X is $A/(A + B)$, where A is the number of persons with grade $(X-1)$ completed who were enrolled in school last year and are not currently enrolled and B is the number of persons with grade X completed who were enrolled last year and are currently enrolled. In computing the rate for the 12th grade, a modification is necessary, since many persons who successfully complete grade 12 will not be enrolled in the fall following graduation. In this case the value for B is the number of persons who were enrolled in the previous fall and who graduated in the spring (as determined from a question that asks high school graduates for their year of graduation.)

Such rates can be ascertained each year from the October Current Population Survey, and the series can easily be extended back in time.[20] Among rates that are available annually and for major population subgroups, this comes closest to recognizing that high school completion is a process that may involve repeated moves out of and back into school. Another important advantage of the annual dropout rates is that they condition on prior school enrollment. Thus, unlike "status" measures of dropout, they are not directly affected by the presence of immigrants who have had no exposure to schooling in the United States.

At the same time, the definition of the annual dropout rate is less than ideal because it combines persons who do not continue from one grade to the next in the survey year with persons who drop out from the next higher grade level during the academic year preceding the survey, as if they were in the same cohort. It also fails to identify return enrollees among this year's students at each grade level. Despite these problems, the definition is useful, perhaps more so than definitions based upon grade completion and enrollment by a specific age, which fail to take account of variation in age-grade progression.[21]

Perhaps to increase its reliability as well as to limit the number of data series that need be displayed, the annual dropout rates are usually aggregated across grades 10 to 12. This also partly overcomes the conceptual problem in cohort coverage mentioned above. For example, the rates are displayed only as aggregated over grades 10 to 12 in the *1994 Green Book* (Committee on Ways and Means, U.S. House of Representatives, 1994: 1142) and in *The Condition of Education 1994* (NCES 1994: 32-33, 359). The rates are broken down by grade level, but only for the total population, in *Dropout Rates in the United States: 1993* (McMillen, Kaufman, and Whitener 1994:3-15) and in the 1992 Current Population Report on School Enrollment (Kominski and Adams 1993). However, the annual dropout rates do not appear either in *Youth Indicators 1993* (NCES 1993a) or in the *Digest of Education Statistics 1993* (NCES 1993b), which do present "status" rates of high school completion.

Because the construction of annual dropout rates by the Bureau of the Census has, since 1992, rested on the official distinction between "12th grade no diploma" and "high school graduate (or equivalent)," there has been a substantial change in the annual rate of high school dropout in the 12th grade. This series, originally presented by McMillen, Kaufman, and Whitener (1994: 13) is reproduced in Figure 14. Obviously, the changing treatment of 12th grade dropout is a major break in the series, and analysts should be most cautious in using the published series.[22] While one may accept or reject the new Census definition of high school completion, there would appear to be a conceptual inconsistency between the definitions of grade completion at the 10th and 11th grade levels, which remain purely nominal, with that at the 12th grade level, which now excludes persons who did not earn a high school diploma or equivalent. As the unit record data from October Current Population Surveys are released publicly, it would be useful, at least for the next few years, to construct an alternative series that will hew more closely to the old definition of 12th grade completion, that is, by including "12th grade no diploma" with "high school graduate."[23]

---

[20] For example, these are the rates shown in Figures 6 to 8 for years since 1972, aggregated across grades 10 to 12, but expressed inversely as rates of persistence, rather than dropout. The series can be extended back to 1967, except in the case of Hispanics.

[21] For further discussion of the conceptualization and measurement of high school dropout, see Kominski (1990) and Pallas (1989).

[22] There is also a minor break in the series between 1986 and 1987, when new editing rules were adopted. The effect of the change was to reduce dropout rates by about 0.5 percentage points.

An important advantage of the annual dropout measure, at least as defined for persons 19 and younger, is that dropout status can be linked to many other characteristics of the household in which the student lives, which is -- at those ages -- almost always a parental or quasi-parental household. Thus, for example, the Census Bureau (Kominski and Adams 1993: Table 8) annually reports dropout rates of dependent family members 15 to 24 years old by age, sex, race, Hispanic origin, and family income. Hauser, Jordan, and Dixon (1993, documented in Hauser and Hauser 1993) have created a uniform file of October Current Population Surveys from 1968 to 1991 in which the characteristics of youths are matched with those of their households, including the social and economic characteristics of the householders.[24] Hauser and Phang (1993) report a logistic regression analysis of trends in high school dropout based on the uniform October files in which family background factors include sex, age, grade level, dependency status, metropolitan locations, region, sex of householder, number of children in the household, educational attainment of householder, occupational status of householder, family income, and housing tenure. These data make it possible to describe the changing social and economic composition of high school cohorts as well as the implications of those changes for high school completion. Over the period 1968 to 1990, the uniform CPS file includes approximately 95,000 whites, 15,000 blacks, and 6,400 Hispanics. Thus, the October CPS data invite multivariate analysis, and the list of regressors is far richer than those used in official statistical series.[25]

Again, the link between dropout or continuation and characteristics of the (parental) household is the presumption that the student lives with parents or parent surrogates. Hauser and Phang (1993) find that fewer than 5 percent of youths are nondependent at the tenth-grade transition; fewer than 10 percent are nondependent at the eleventh-grade transition; and fewer than 15 percent are nondependent at the twelfth grade transition. In general, nondependency is greater among women than men, and it is greater among Hispanics than among whites or blacks. The older the person, and the higher the grade level, the less likely that he or she is dependent.

When an individual is living independently, household income may well be an effect, rather than a source of high school dropout, continuation, or completion. So far as high school completion is concerned, I think that the link between children and parental households should be regarded as questionable no later than age 20, and perhaps by age 19. Thus, in my opinion, it would not be appropriate to link records of high school completion by ages 21-22 (or older) to characteristics of youth's households on the same assumption that apply to the dropout rate at the 10th to 12th grade levels. I am highly skeptical of the suggestion by McMillen, Kaufman, and Whitener (1994: 130-31) that the inverse relationship between dropout and income, regardless of dependency status, shows there is no problem in arraying "status" dropout rates at ages 15 to 24 by household income. Having accepted this notion, it would be but a small step to array dropout rates among nondependents by their completed levels of schooling.

---

[23] The effect of the changing definition is especially large among overage students covered by the annual dropout concept, that is, persons aged 20 to 24, and there is scarcely a blip in the series below age 20. Thus, an alternative to revising the definition of high school completion used in the series would be to limit the dropout rate to students aged 15 to 19.

[24] This file is available from the Inter-university Consortium for Political and Social Research.

[25] In *Dropout Rates in the United States: 1993* marginal event dropout and persistence rates are presented by sex, race-ethnicity, family income, region, and metropolitan status. Time series of rates are presented by race-ethnicity by sex, by grade level, and by age.

The major limitation of the annual dropout (or persistence) rates is the obverse of their most attractive feature. They pertain to dropout or completion or a single grade in a single year or to dropout or completion aggregated across three grade levels in a single year. Thus, as suggested in some of the series presented earlier, the data are rather thin, and simple disaggregations of rates, e.g., those presented in official publications, are subject to a great deal of sampling variability. On the other hand, if one is willing to think of year as a variable, rather than as an obligatory unit for aggregation, analysis, and reporting, or if one is willing to assume relative constancy in the effects of some social and economic characteristics across limited periods of time, e.g., by aggregating or smoothing data across years, we can readily obtain a far richer understanding of the sources of trends and differentials in dropout.

For example, Figure 15 shows estimated annual high school dropout rates by grade level, sex, and race-ethnicity from 1973 to 1989.[26] The estimates are based upon a logistic regression model that includes main effects on dropout of grade level, sex, and race-ethnicity; interaction effects between grade level and sex, race-ethnicity and sex, and race-ethnicity and grade level; and interaction effects between year and grade level and between year and race-ethnicity. This model has the effect of smoothing the data, for it does not include the three-way interaction effects of dropout with grade level, sex, and race-ethnicity or any of the higher-order interaction effects of dropout with year and grade level, sex, or race-ethnicity. All of these higher-order interaction effects were tested and found not to be statistically significant. Thus, the model estimates distinct trends in dropout by grade level and race-ethnicity, but not by sex, and the trends for combinations of grade level and race-ethnicity are combinations of the trends for each grade level and each race-ethnicity. The model increases statistical power, that is, it decreases standard errors, for the significant contrasts.

The figure shows the trends in annual high school dropout rates by race-ethnicity within each of six combinations of sex and grade level.[27] Dropout rates are consistently higher with each successive grade level and, among blacks and Hispanics, they are substantially higher at the twelfth grade level than at the tenth or eleventh grades. Among whites at all grade levels there was a steady decline in high school dropout during the 1980s. Among blacks, dropout rates were generally lower in the 1980s than in the 1970s. The trends are less clear among Hispanics, but the data suggest an increase in dropout through the early to middle 1980s, followed by a decline to the level of the middle 1970s. In each combination of grade level and sex, rates of dropout are almost always highest among Hispanics, followed by African Americans; dropout rates are lowest among whites. However, black and Hispanic dropout rates are similarly high at the twelfth-grade level, where the gap between the minority groups and whites is larger than in the tenth or eleventh grade. Fewer than 10 percent of white men or women have dropped out of school in any year or grade level since the early 1970s. Among blacks, fewer than 10 percent drop out of school at the tenth or eleventh grade, but about 15 percent drop out in the twelfth grade. Among Hispanics, about 10 percent drop out in the tenth and eleventh grade, but 15 percent or more of men and 10 to 15 percent of women drop out in the twelfth grade. At the twelfth-grade level, there is also a higher rate of dropout among men than among women in each racial-ethnic group.

In Figure 16, under similar assumptions about interactions between race-ethnicity, sex, grade-level, and year, the annual dropout rates have been adjusted for effects of social background, both within and between years.[28] Thus, the model permits comparisons of dropout rate across years, within and among race-ethnic groups. The rates have been normed in two ways. First, they pertain to dropout among dependent youth, not to all high school students.[29] Second, rates of dropout are normed so the predicted rates among dependent black youth (of each sex and at each grade level) are set equal to the corresponding observed rates. By virtue of this normalization, the dropout rates of whites and of Hispanics can be said to pertain to youth in those groups with the average social background characteristics of blacks.

---

[26] These are reproduced from Hauser and Phang (1993).

[27] The six panels of this and the next figure are prepared to the same scale in order to facilitate comparison.

[28] Social background includes all of the variables mentioned earlier that are available in the uniform October CPS files.

[29] This is a norming assumption; it does not mean that the model pertains only to dependent youth. Data for nondependent youth have been used to estimate effects of dependency status, sex, year, grade level, race-ethnicity, and regional and metropolitan location.

The striking finding in figure 16, which does not depend on the normalization of the dropout rates, is that controls for social background reverse the observed ordering of dropout rates between whites and blacks or Hispanics, especially in the 1970s. That is, when social background is controlled, whites have the highest propensity to drop out of high school, followed by Hispanics and then by blacks. Moreover, by the end of the 1980s, and primarily because of a steady improvement among whites, there was a substantial convergence in dropout rates among the three racial-ethnic groups. That is, one need not invoke either culture, motivation, or discrimination to account for observed racial-ethnic differences in high school dropout; they are fully explained by easily observable factors of social and economic origin. Moreover, to the degree that we must offer some explanation of differences in dropout beyond the obvious, the problem is not to explain higher dropout among minorities, but to explain it in the majority population. One possibility is that economic opportunities outside of school were greater for whites than for minorities. This would account both for the higher net rates of dropout among whites and -- because of the global decline in the labor market opportunities of dropouts -- for the convergence of white dropout rates with those among minorities.

There are real limits to our ability to disaggregate annual dropout rates. For example, I have not yet been successful in constructing useful time series of rates from the CPS data by state or for major metropolitan areas.[30] However, I believe that further analytic work, using status rates, say, for 20 to 24 year olds, as well temporal aggregations of annual dropout rates, may prove fruitful.

Cohort Rates

The NCES dropout reports suggest a distinction among "event," "status," and "cohort" rates. They use the latter term primarily to refer to rates of dropout or high school completion in the major longitudinal studies carried out by the NCES, that is, the High School and Beyond surveys of the early 1980s (HSB) and the National Educational Longitudinal Study of 1988 (NELS:88). Generically, NCES says that, "Longitudinal or cohort analyses are based on repeated measures of a group of individuals." In *Dropout Rates in the United States: 1993*, cohort rates are illustrated by comparison of dropout status in repeated CPS cross-sections, as well as by reference to the longitudinal studies (McMillen, Kaufman, and Whitener 1994: 32-33). In my opinion, the distinction between "cohort" and "event" or "status" rates has not been cleanly drawn, nor has the NCES used the available longitudinal data in its annual reports as well or as thoroughly as one might hope.

For example, the annual dropout rates discussed in the preceding section would qualify as "cohort" rates, as those are described in the NCES reports, because they depend upon longitudinal observation. To be sure, the initial condition -- enrollment in the prior year -- is ascertained retrospectively, but that does not alter the concept. Thus, there is little difference in concept between annual dropout rates from the CPS and the 8th to 12th or 10th to 12th grade dropout rates from NELS, except the latter are based upon a longer period of observation. This is not to say that reported findings (and comparisons of findings) from HSB and NELS:88 are unimportant, but merely that they do not offer any conceptual advantage relative to the CPS. The difference seems smaller yet when one considers the large array of social background variables that are available in the CPS, but not used in the dropout reports, as well as the richer set of variables available in the longitudinal studies that have not yet been used in the official reports.

One area in which the longitudinal studies could be most valuable, and the NCES reports are incomplete -- one might even say that they are timid -- is the relationship between academic performance and school dropout. The National Educational Goals call on us to improve academic performance as well as school retention. There is an obvious relationship between academic success and retention, which is glossed through references to motivation, effort, and engagement of students (OERI 1993: 2). There may well be trade-offs between growth in academic standards and in school retention. Unlike the Current Population Survey, the major national longitudinal studies do contain good measures of academic achievement. Yet the NCES dropout reports contain almost no information about the relationship between academic performance and school retention or dropout. In *Dropout in the United States: 1993*, dropout from NELS:88 is reported by sex, race-ethnicity, self-reported reasons for dropping out, poverty level, family composition, and presence of an own child in the home. The report presents no direct evidence about the relationship between academic performance or grade retention and school dropout. In my opinion, this is a deplorable omission.

---

[30] The October CPS data are reported separately and consistently for 17 very large metropolitan areas from 1968 to the present. It is not possible, for reasons of confidentiality, to identify all states, but most states are identifiable in each survey year.

I have elsewhere suggested that the value of the national longitudinal studies could be increased substantially, at relatively low cost and without sacrificing analytic utility, if the observations were spread out over the decade on an annual or biennial basis (Hauser 1991). I have seen no progress toward this in the past few years, but I believe that it remains an attractive goal. Briefly, by spreading observations across calendar years, rather than "bunching" them in cohort samples drawn once per decade, we can accumulate samples of analytic utility equal to those presently available. But at the same time, we could obtain readings of trend in the process of schooling on a regular basis, rather than once per decade. A redesign of the national longitudinal surveys should also include oversamples of large, minority populations that will be large enough to monitor trends within those groups, as well as among non-Hispanic whites or the total population. Thus, improvements in the design, as well as in the use of longitudinal educational data could contribute to educational policy with respect to high school dropout and completion.

Future Prospects

In closing, I want to comment briefly on two other ways in which data on school dropout and retention may be measured in future. The first is the use of administrative records. An ambitious effort is now underway to obtain comparable measures of high school dropout across states and across time through the Common Core of Data (CCD), an annual survey of state-level educational agencies (McMillen, Kaufman, and Whitener 1994: 61-63). During 1991-92, an effort was made by 43 states to participate in this new program, but only 15 states reported data that were consistent with the specified definitions of school enrollment and dropout. Moreover, these administrative data depend on the ability of local and state agencies to determine whether or not a student who has left a school has subsequently re-enrolled elsewhere. The goal of the program is to report "the number and rate of event dropouts from public schools by school districts, states, major subpopulations, and the nation ... by grade for grades 7-12, by sex and by sex within race-ethnicity categories" (p. 63). This is an ambitious and laudable undertaking. If it is successful, it will obviously fill in many of the inter-censal gaps in reports of school dropout and retention for state and local areas. However, given the difficulty of imposing standard definitions across localities and states and of linking school-leavers to re-enrollment, I suspect that it will be a long time before this project yields useful cross-sections or time-series.

The Census Bureau's plans for continuous measurement are another distant, but attractive possibility for the improvement of educational indicators below the national level. Briefly, the Bureau is hoping to introduce a very large national sample survey operation, perhaps with as many as 250,000 households per month. This would eventually replace the long form in the decennial census of population, and it would be designed to produce reliable data for very small areas, e.g., census tracts, when the data are accumulated over a 3 to 5 year period. Assuming that the content of this survey would be similar to that of recent decennial censuses or to the March Current Population Survey, it would present very rich possibilities for estimation of school dropout and completion on a regular basis, well below the national level. Very serious design and operational issues must be addressed before continuous measurement becomes a reality, but it does offer the possibility of vast improvement in our ability to monitor the process of schooling.

REFERENCES

Annie E. Casey Foundation. 1994. Kids Count Data Book: State Profiles of Child Well-Being. Baltimore, Maryland: Annie E. Casey Foundation.

Astone, Nan Marie, and Sara McLanahan. 1991. "Family Structure, Parental Practices, and High School Completion." *American Sociological Review* 56(3)(June): 309-20.

Bielby, William T., Robert M. Hauser, and David L. Featherman. 1977. "Response Errors of Black and Nonblack Males in Models of the Intergenerational Transmission of Socioeconomic Status." *American Journal of Sociology* 82 (May): 1242-88.

Children's Defense Fund. 1994. The State of America's Children Yearbook: 1994. Washington, D.C.: Children's Defense Fund.

Committee on Ways and Means, U.S. House of Representatives. 1994. The Green Book 1994: Background Material and Data on Programs Within the Jurisdiction of the Committee on Ways and Means. Washington, D.C.: GPO.

Ekstrom, Ruth B., Margaret E. Goertz, Judith M. Pollack and Donald A. Rock. 1986. "Who Drops Out of High School and Why? Findings from a National Study." *Teacher's College Record* 87(3)(Spring): 356-73.

Frase, Mary J. 1989. "Dropout Rates in the United States: 1988." *Analysis Report*, NCES 89-609. National Center for Education Statistics. Washington, D.C.: GPO.

Hauser, Robert M. 1991. "Measuring Adolescent Educational Transitions Among Black Americans, Hispanics, and Whites." Institute for Research on Poverty. Discussion Papers, no. 951-91. Madison, Wisconsin: University of Wisconsin-Madison.

-------. 1993. "Trends in College Entry Among Blacks, Hispanics, and Whites." *Studies of Supply and Demand in Higher Education*. Edited by Charles Clotfelter and Michael Rothschild. National Bureau of Economic Research. Chicago, Illinois: University of Chicago Press.

Hauser, Robert M., and Taissa S. Hauser. 1993. *Current Population Survey, October Person-Household Files, 1968-90: Cumulative Codebook*. Principal Investigator, Robert M. Hauser. Madison, Wisconsin: Center for Demography and Ecology, Department of Sociology, University of Wisconsin-Madison.

Hauser, Robert M., and Hanam Samuel Phang. 1993. "Trends in High School Dropout Among White, Black, and Hispanic Youth, 1973 to 1989." IRP Discussion Paper, vol. 1007-93. Madison, Wisconsin: Institute for Research on Poverty.

Hauser, Robert M., Linda Jordan, and James A. Dixon. 1993. *Current Population Survey, October Person-Household Files, 1968-90*. [Machine-readable data files] Principal Investigator, Robert M. Hauser. Madison, Wisconsin: Center for Demography and Ecology, Department of Sociolgy, University of Wisconsin-Madison.

Haveman, Robert, Barbara L. Wolfe, and James Spaulding. 1991. "Educational Achievement and Childhood Events and Circumstances." *Demography* 28(1)(February): 133-57.

Jaeger, David A. 1993. "The New Current Population Survey Education Variable: A Recommendation." Research Reports. 93-289. The University of Michigan: Population Studies Center.

Kaufman, Phillip, and Mary J. Frase. 1990. "Dropout Rates in the United States: 1989." *Analysis Report*, NCES 90-659. National Center for Education Statistics. Washington, D.C.: GPO.

Kaufman, Phillip, Marilyn M. McMillen, and S. Whitener. 1991. "Dropout Rates in the United States: 1990." *Analysis Report*, NCES 91-053. National Center for Education Statistics. Washington, D.C.: GPO.

Kaufman, P., M.M. McMillen, E. Germino-Hausken and D. Bradby. 1992. "Dropout Rates in the United States: 1991." *Analysis Report*, NCES 92-129. National Center for Education Statistics. Washington, D.C.: GPO.

Kominski, Robert. 1990. "Estimating the National High School Dropout Rate." *Demography* 27(2)(May): 303-11.

Kominski, Robert, and Andrea Adams. 1993. "School Enrollment - Social and Economic Characteristics of Students: October 1992." *Current Population Reports*. Population Characteristics. Series P20, vol. 474. Washington, D.C.: U.S. Bureau of the Census October.

-------. 1994. "Educational Attainment in the United States: March 1993 and 1992." *Current Population Reports*. Population Characteristics. Series P20, vol. 476. Washington, D.C.: U.S. Bureau of the Census May.

Krein, Shelia Fitzgerald, and Andrea H. Beller. 1988. "Educational Attainment of Children From Single-Parent Families: Differences by Exposure, Gender, and Race." *Demography* 25(2)(May): 221-34.

Mare, Robert D. 1995. "Changes in Educational Attainment, School Enrollment, and Skill Levels." In *Economic Trends*. Vol. I. Edited by Reynolds D. Farley. State of the Union: America in the 1990s. New York: Russell Sage Foundation.

McLanahan, Sara. 1985. "Family Structure and the Reproduction of Poverty." *American Journal of Sociology* 90(4): 873-901.

McMillen, M.M., P. Kaufman, and S. Whitener. 1994. "Dropout Rates in the United States: 1993." *Analysis Report*, NCES 94-669. National Center for Education Statistics. Washington, D.C.: GPO.

McMillen, M.M., P. Kaufman, E. Germino-Hausken and D. Bradby. 1993. "Dropout Rates in the United States: 1992." *Analysis Report*, NCES 93-464. National Center for Education Statistics. Washington, D.C.: GPO.

Murnane, Richard J., and Frank Levy. 1993. "Why Today's High-School Educated Males Earn Less Than Their Fathers Did: The Problem and an Assessment of Responses." *Harvard Educational Review* 63(1)(Spring): 1-20.

Murphy, Kevin, and Finis Welch. 1989. "Wage Premiums for College Graduates: Recent Growth and Possible Explanations." *Educational Researcher* 18(4)(May): 17-26.

National Center for Education Statistics. 1994. The Condition of Education 1994, NCES 94-149. Washington, D.C.: GPO.

The National Educational Goals Report: Building a Nation of Learners National Educational Goals Panel. 1994. Washington, D.C., September.

National Center for Education Statistics. 1993a. Youth Indicators 1993, NCES 93-242. Washington, D.C.: GPO.

-------. 1993b. Digest of Education Statistics 1993, NCES 93-292. Washington, D.C.: GPO.

Office of Educational Research and Improvement. 1993. Reaching the Goals: Goal 2, High School Completion, Technical Report. Washington, D.C.: U.S. Department of Education December.

Pallas, Aaron M. 1989. "Conceptual and Measurement Issues in the Study of School Dropouts." Pp 87-116 in *Research in the Sociology of Education and Socialization*. Greenwich, Connecticut: JAI Press.

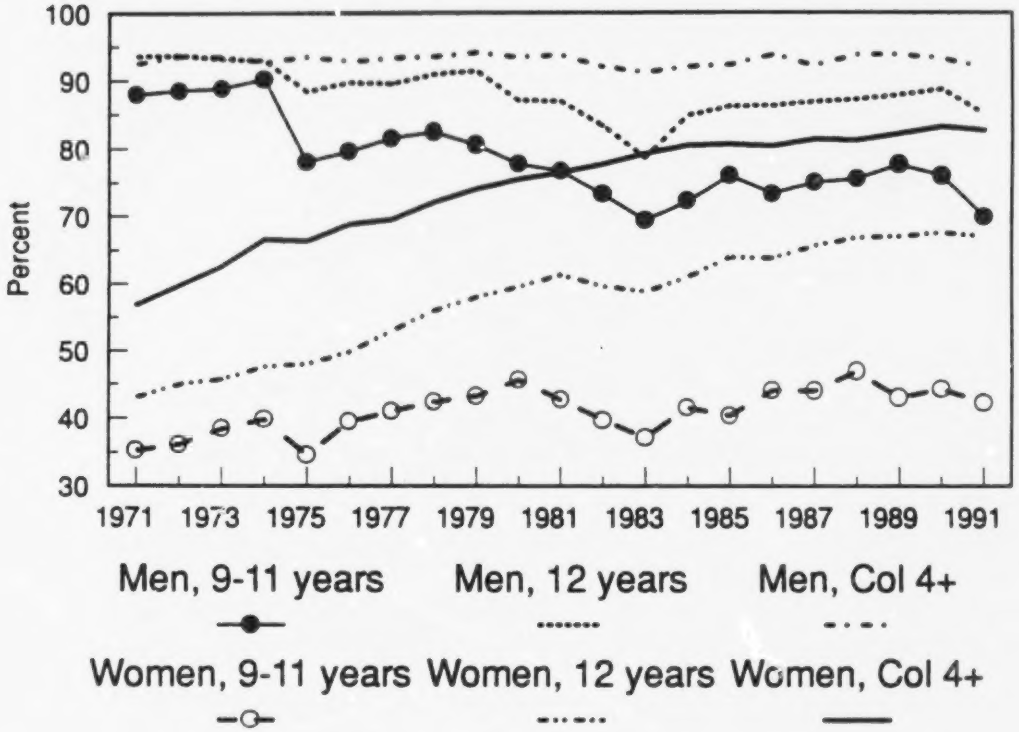
Sandefur, Gary D., Sara McLanahan, and Roger A. Wojtkiewicz. 1992. "The Effects of Parental Marital Status During Adolescence on High School Graduation." *Social Forces* 71(1)(September): 103-21.

Schrieber, Daniel (ed). 1967. *Profile of the School Dropout*. New York: Vintage.

U.S. Bureau of the Census. 1994. Education in the United States, 1990 CP-3-4. 1990 Census of Population. Washington, D.C.: GPO.

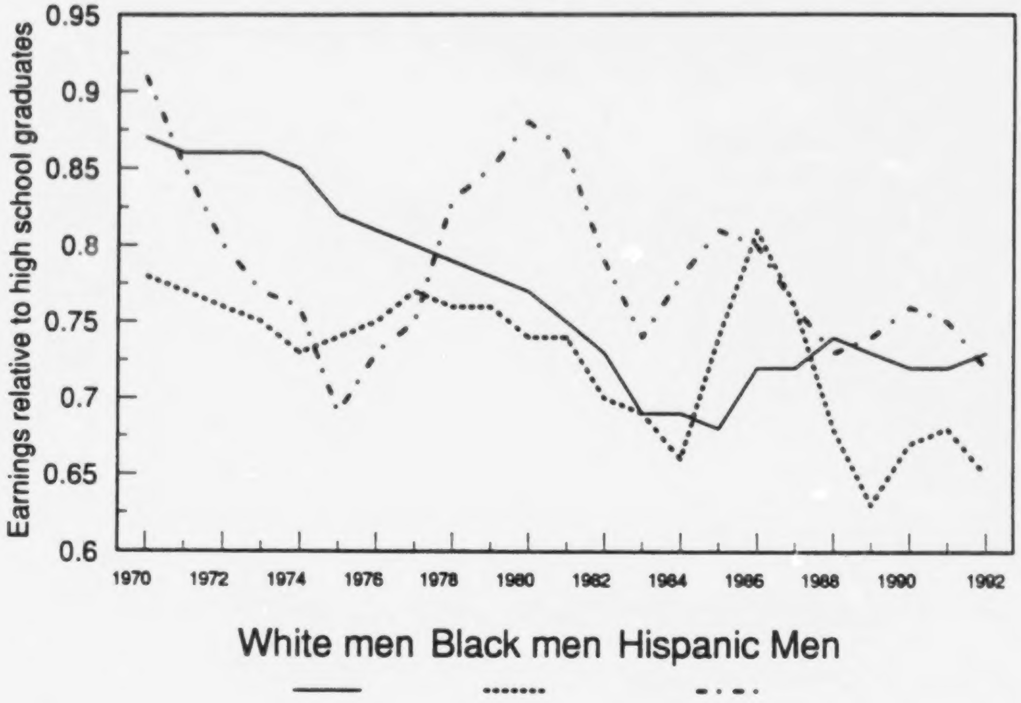
U.S. Department of Education. 1990. "National Goals for Education." Washington, D.C.: GPO.

Figure 1. Employment Rates of Persons 25 to 34 Years Old by Sex and Educational Attainment, 1971 to 1991



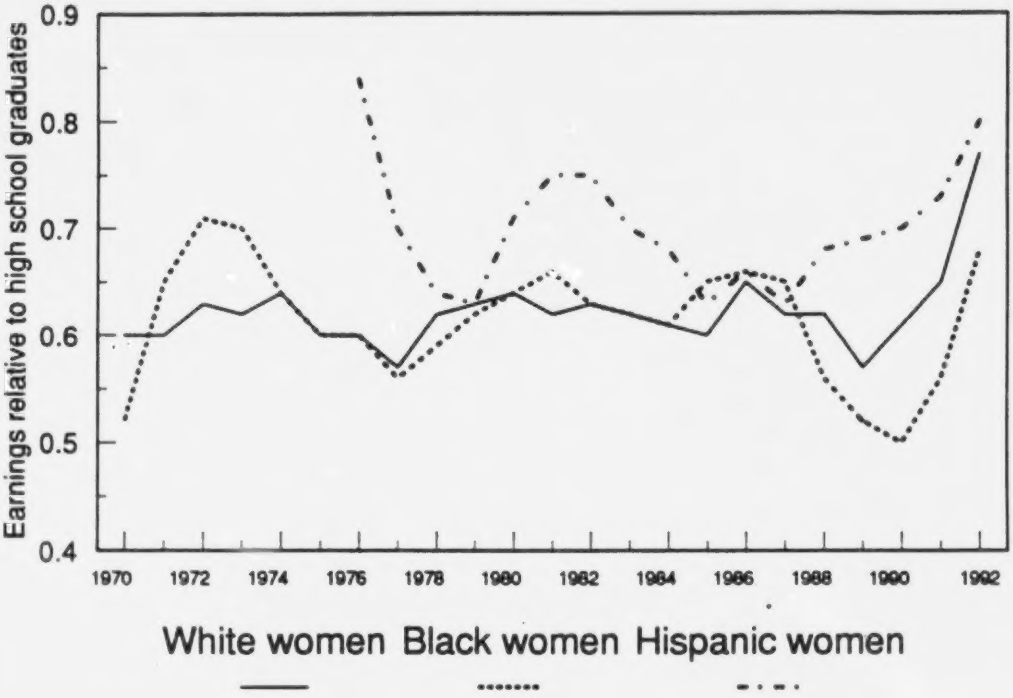
Source: Current Population Survey

Figure 2. Ratio of Median Annual Earnings of 25 to 34 Year Old Male Wage and Salary Workers with 9-11 Years of School to those with 12 Years of Schooling by Race/Ethnicity: 1970 to 1992



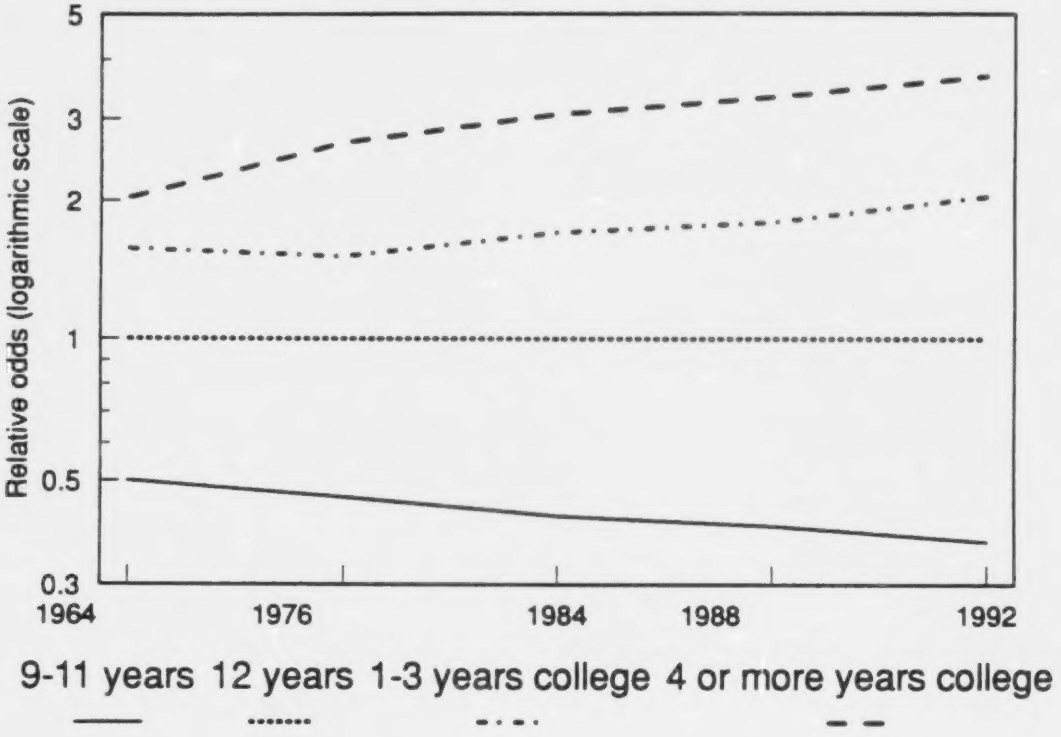
Note: Data are 3-year averages from March Current Population Surveys.

Figure 3. Ratio of Median Annual Earnings of 25 to 34 Year Old Female Wage and Salary Workers with 9-11 Years of School to those with 12 Years of Schooling by Race/Ethnicity: 1970 to 1992



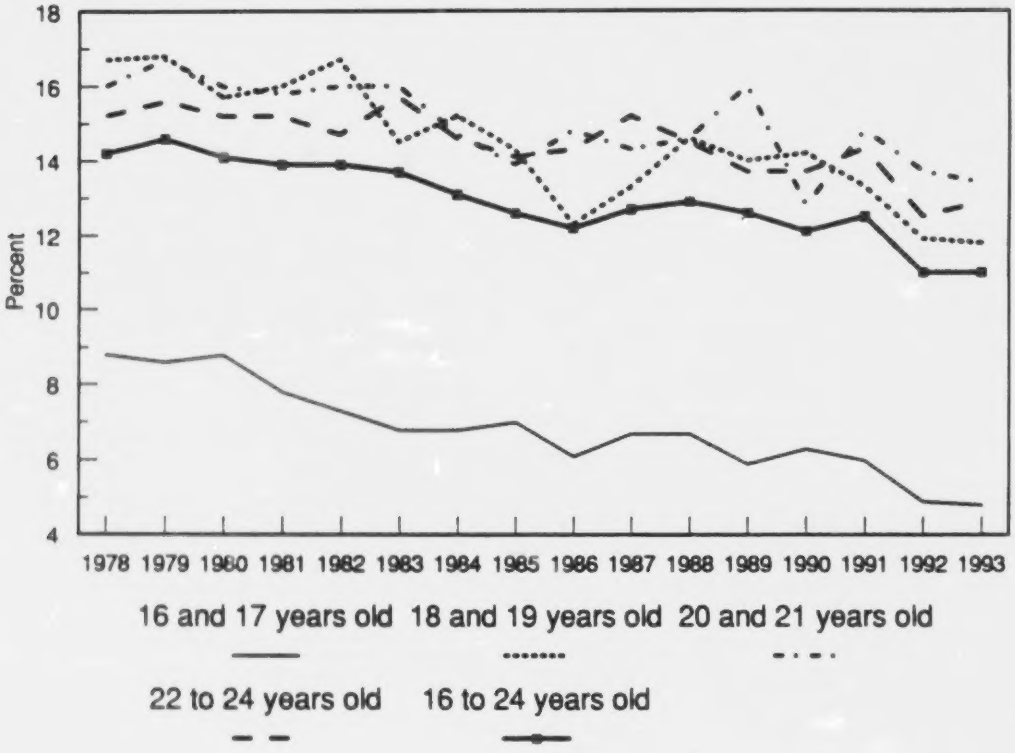
Note: Data are 3-year averages from March Current Population Surveys.

Figure 4. Odds of Voting in Selected Presidential Elections, Relative to the Odds among High School Graduates



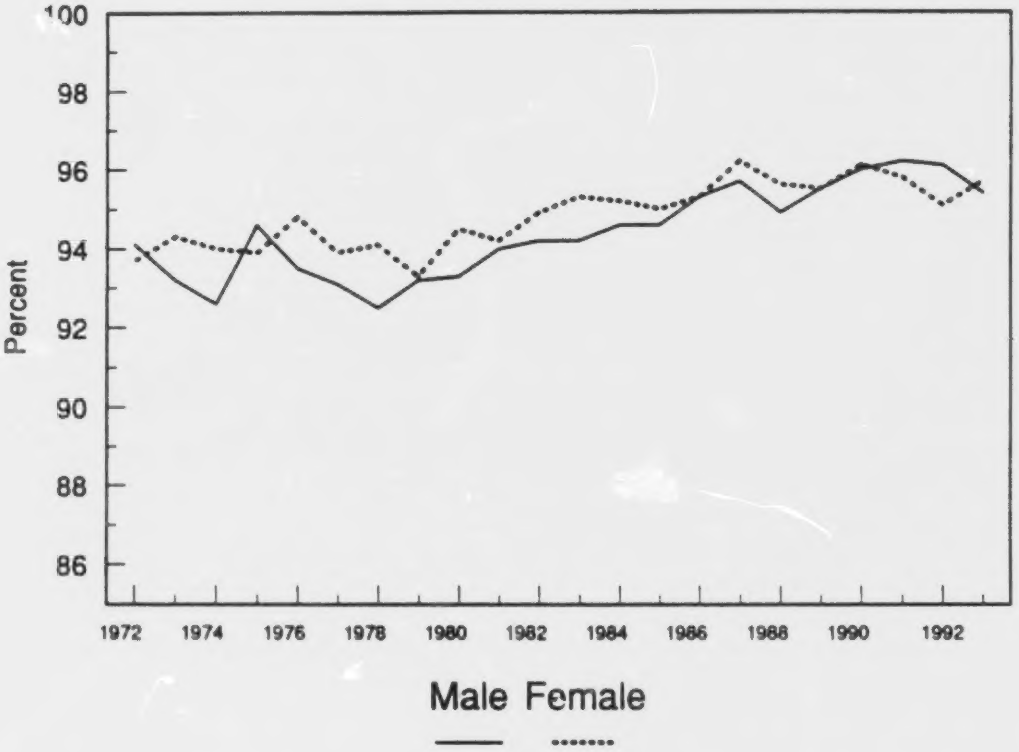
Source: Current Population Surveys

Figure 5. Persons Who Were Not Enrolled in School and Had Not Graduated from High School by Age, 1978 to 1993



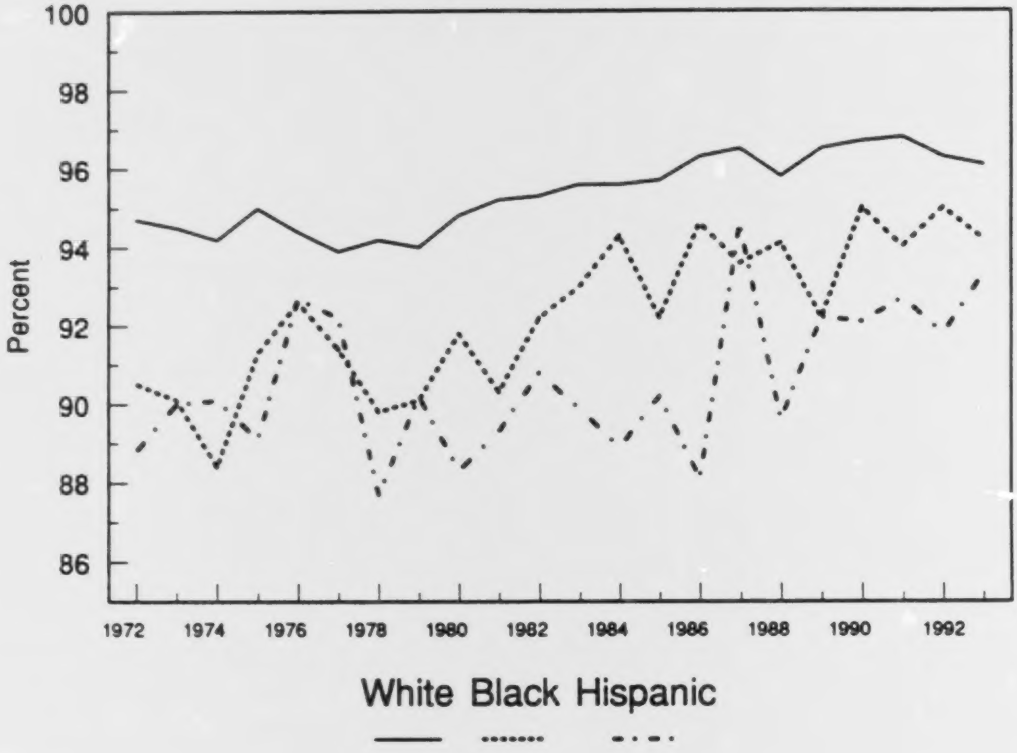
Source: October Current Population Surveys

Figure 6. Year-to-Year Persistence of High School Students in Grades 10-12, Ages 15-24, by Sex, 1972-1993



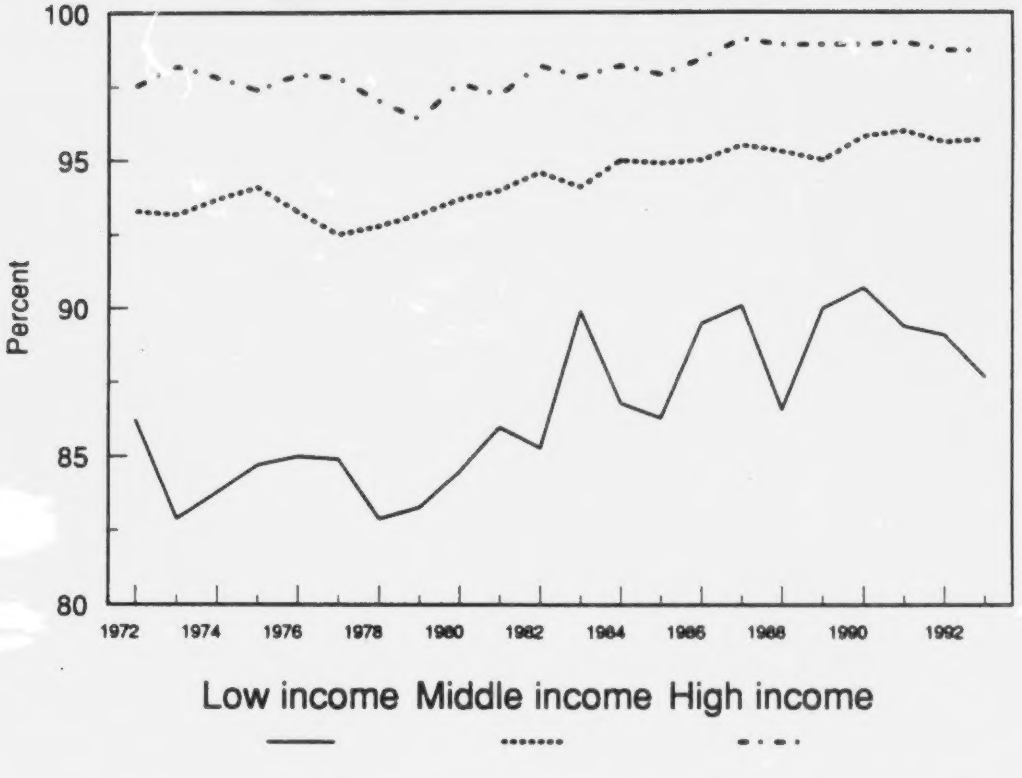
Source: October Current Population Surveys

Figure 7: Year-to-Year Persistence of High School Students in Grades 10-12, Ages 15-24, by Race-Ethnicity: 1972-93



Source: October Current Population Surveys

Figure 8. Year-to-Year Persistence of High School Students in Grades 10-12, Ages 15-24, by Family Income, 1972-1993



Source: October Current Population Surveys

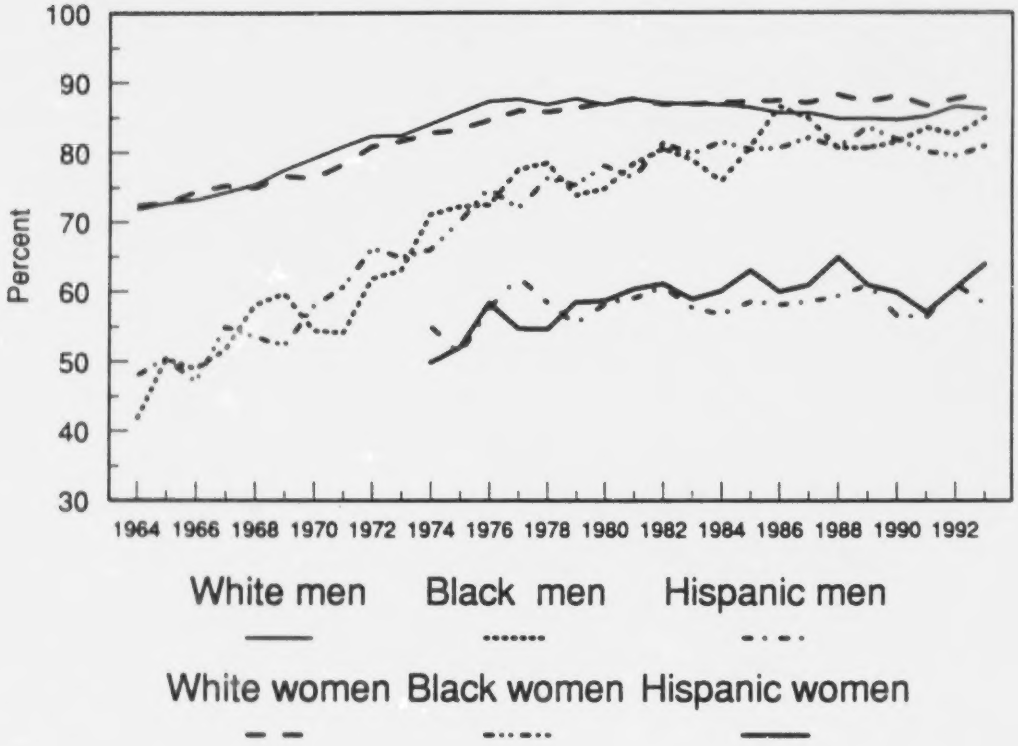


Figure 9. Percentage of Persons Aged 25 to 29 Completing 12 or More Years of Schooling, by Race/Ethnicity and Sex: 1964 to 1993

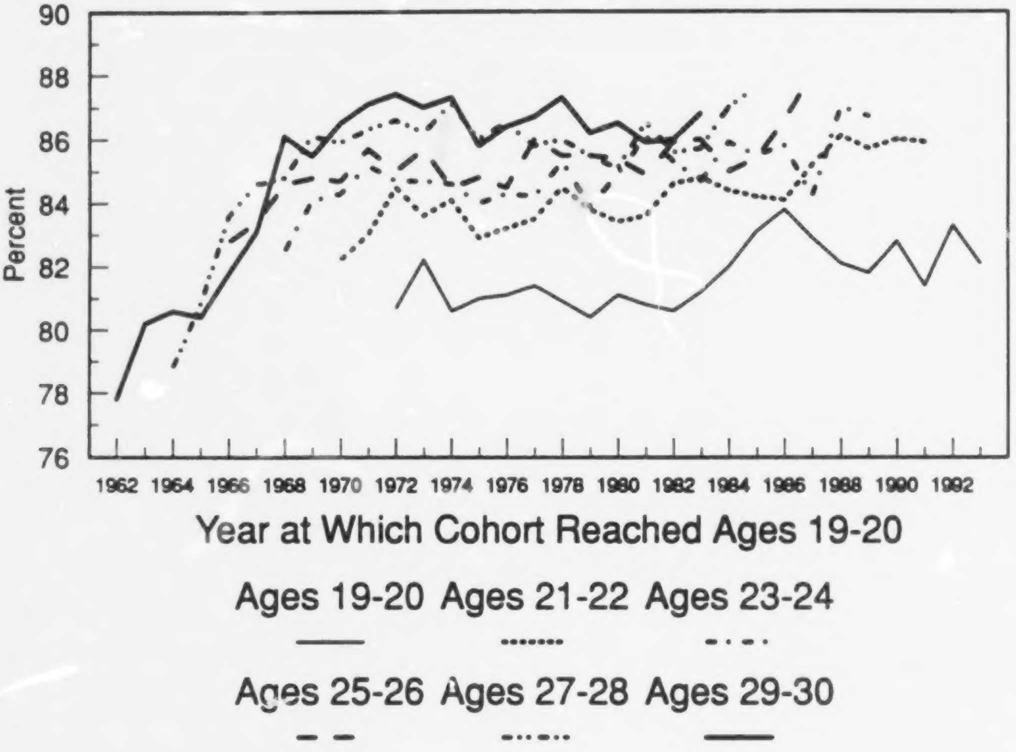
Source: March Current Population Surveys

Figure 10. The 1990-Basis CPS Educational Attainment Classification

What is the highest level of school ... has completed or the highest degree ... has received?

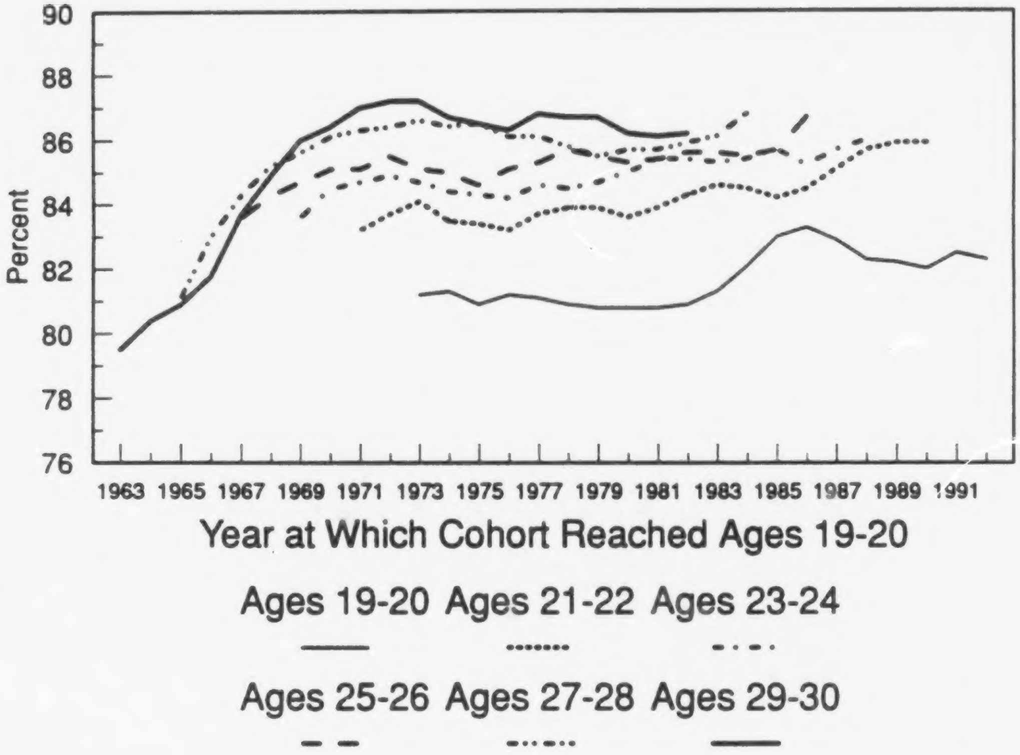
Code	Level of Schooling Completed
31	Less than first grade
32	1st, 2nd, 3rd, or 4th grade
33	5th or 6th grade
34	7th or 8th grade
35	9th grade
36	10th grade
37	11th grade
38	12th grade NO DIPLOMA
39	HIGH SCHOOL GRADUATE - high school diploma or the equivalent (For example, GED)
40	Some college but no degree
41	Associate degree in college - Occupational/vocational program
42	Associate degree in college - Academic program
43	Bachelor's degree (For example: BA, AB, BS)
44	Master's degree (For example: MA, MS, MEng, MEd, MSW, MBA)
45	Professional School Degree (For example: MD, DDS, DVM, LLB, JD)
46	Doctorate degree (For example: PhD, EdD)

Figure 11. High School Completion of Selected Cohorts by Ages 19 to 20, 21 to 22, 23 to 24, 25 to 26, 27 to 28, and 29 to 30: 1962 to 1993



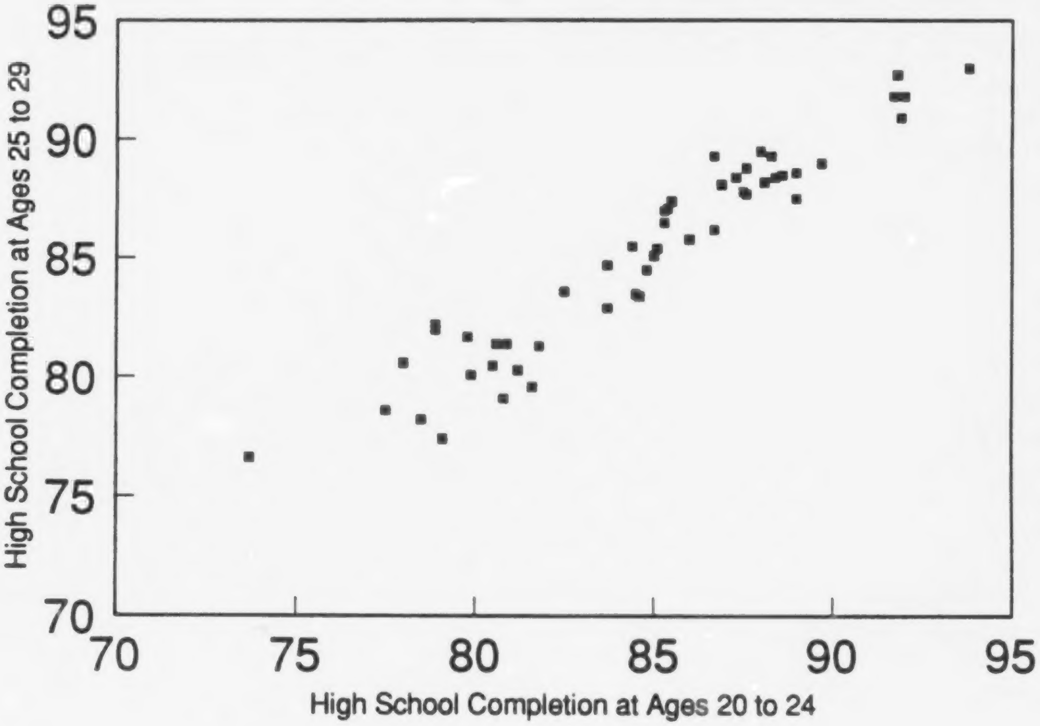
Source: March Current Population Surveys

Figure 12. Smoothed Estimates of High School Completion of Selected Cohorts by 19 to 20, 21 to 22, 23 to 24, 25 to 26, 27 to 28, and 29 to 30: 1963 to 1992



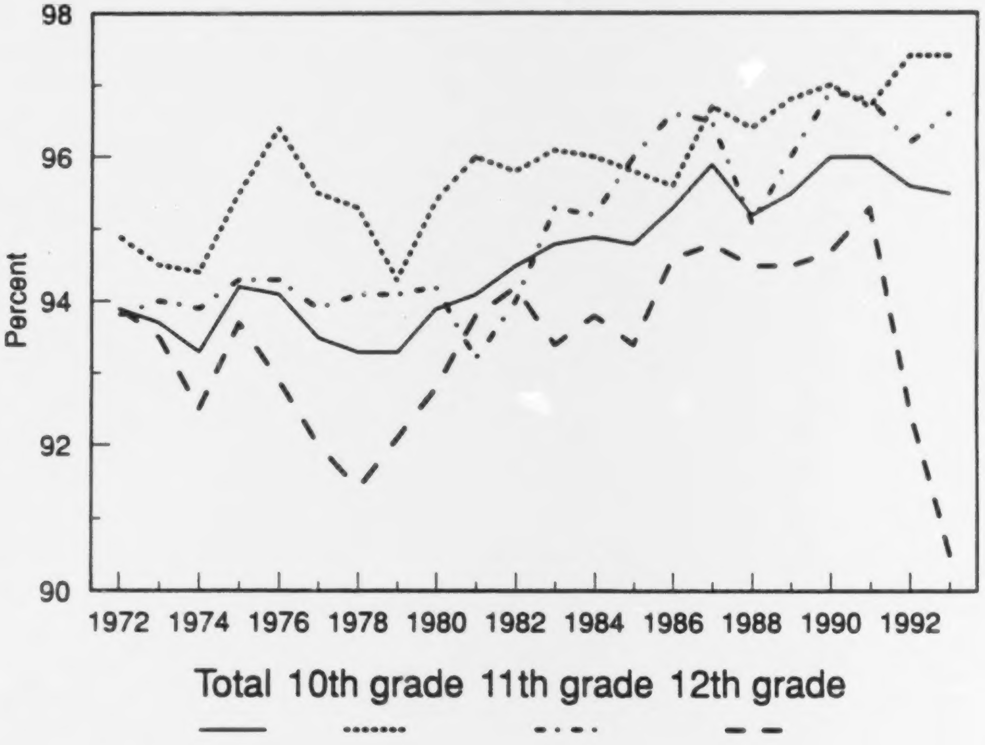
Source: March Current Population Surveys (3-year averages)

Figure 13. Percentage Completing High School at Ages 20 to 24 and 25 to 29 by States: April 1990



Source: 1990 Census of Population

Figure 14. Year-to-Year Persistence of High School Students, Aged 15 to 24 in Grades 10 to 12 by Grade Level, 1972 to 1993

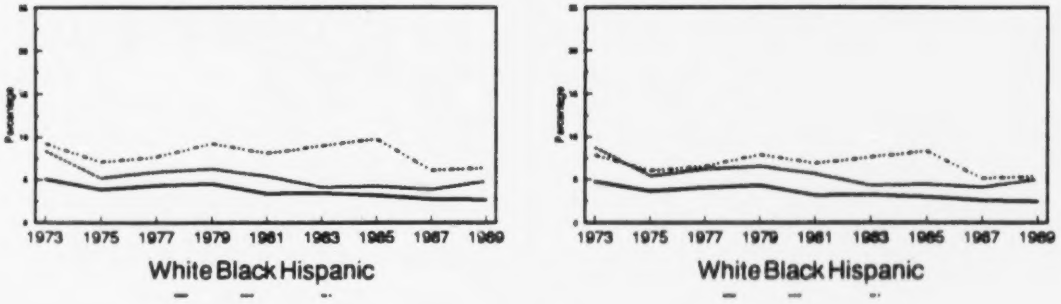


Source: October Current Population Surveys

Figure 15. Annual High School Dropout by Race-Ethnicity: Tenth to Twelfth Grade Males and Females, 1973 to 1989

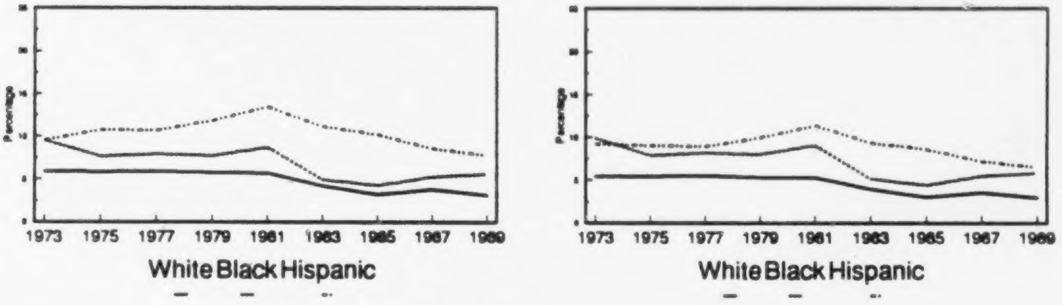
Tenth Grade Males

Tenth Grade Females



Eleventh Grade Males

Eleventh Grade Females



Twelfth Grade Males

Twelfth Grade Females

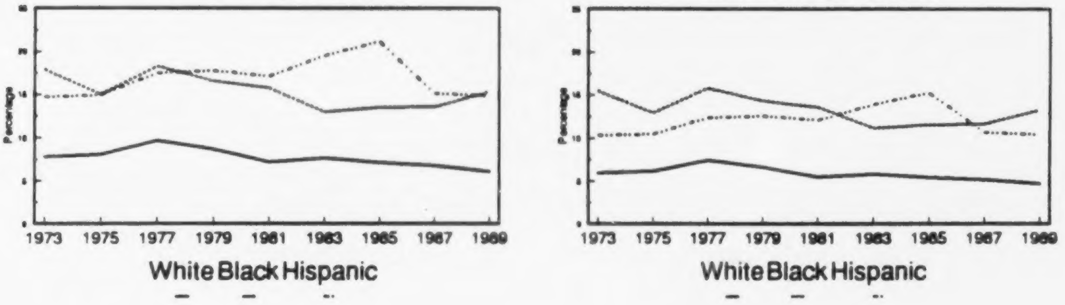
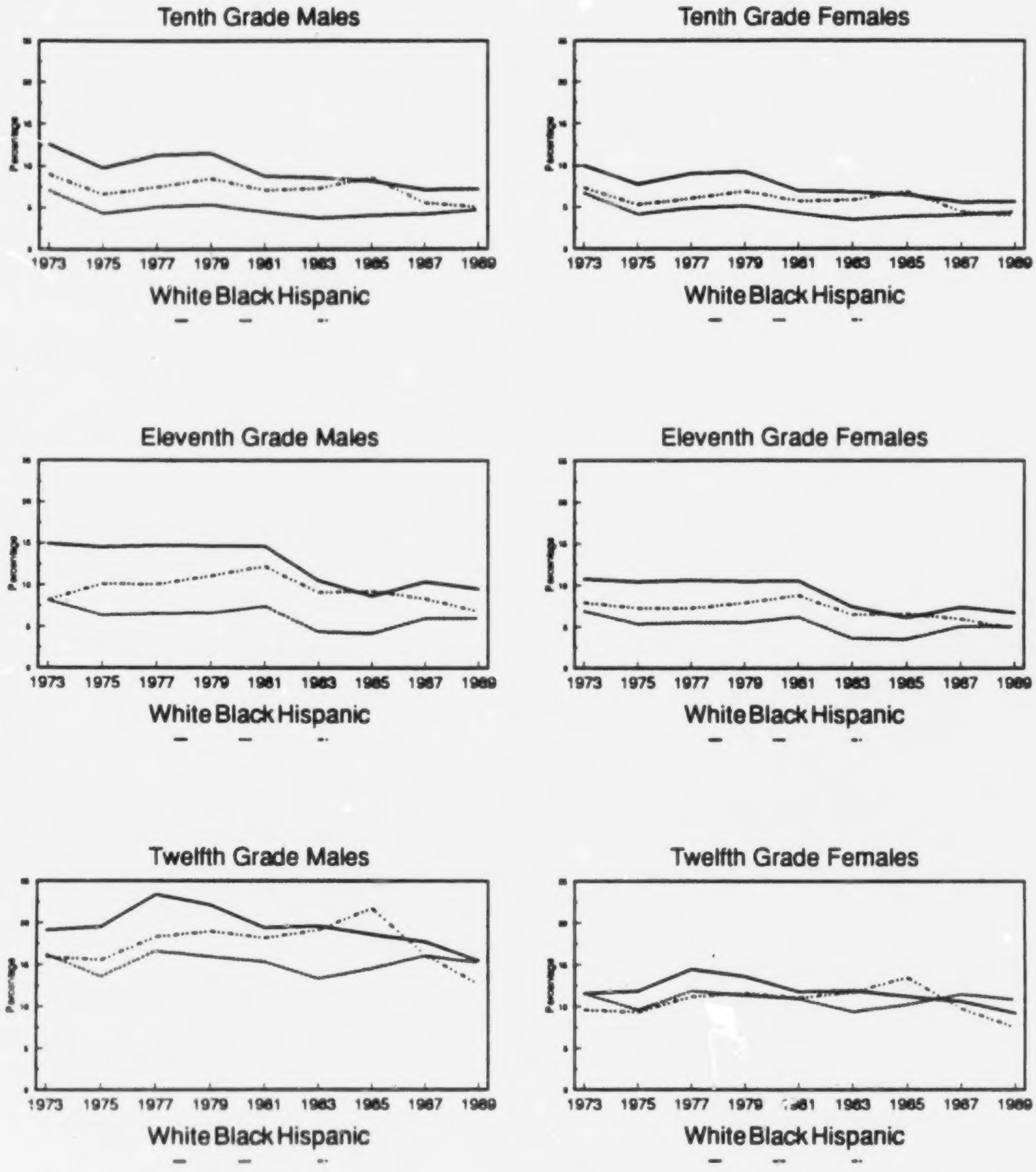


Figure 16. High School Dropout Adjusted for Social Background:
Tenth to Twelfth Grade Men and Women by Race-Ethnicity, 1973 to 1989



Note: All graphs are plotted for dependents with the characteristics of average black males or females at each grade level.

Postsecondary and Vocational Education:
Keeping Track of the College Track

by Thomas J. Kane
Brookings Institution and
Harvard University

October 21, 1994

This paper was prepared for the conference on "Indicators of Children's Well-Being", November 17-18, 1994. Anthony Shen provided research assistance.

I. Introduction

Throughout the Eighties, the value of a college education increased dramatically as the earnings prospects of high school graduates dimmed. As a result, the stakes have been raised in the debate over college costs, access to college and the payoffs to different types of postsecondary education. While high school graduation was the critical hurdle facing youth two decades ago, college attendance is increasingly the prerequisite for a decent standard of living today. Unfortunately, our data collection methods have failed to keep pace with these important developments in the labor market, leaving us guessing about many crucial questions regarding the well-being of youth of college-age. In the following chapter, I survey the most important measures related to postsecondary education. The discussion is organized around three primary areas of concern: access (Who is enrolled in postsecondary education and in various types of institutions?), cost (How much are students and the state and federal government investing in postsecondary education?) and the payoffs to different types of education (How much does such an education seem to influence one's earning prospects later in life?). Each section contains a description of available data and suggestions for revising available measures.

The four primary recommendations are described below:

o Collect parental education and occupation information for young adults (16-24) on the Current Population Survey (CPS).

Given the CPS household definition, it is not possible to match youth outcomes to parental characteristics once they move out of the household. As a result, though real public tuition levels increased by 50% during the Eighties, the impacts on the gap in college entry for youth of varying socioeconomic backgrounds is unclear. Given the rising importance of college entrance one's lifetime earnings prospects, this hole in our statistical system should be filled.

o Develop a small number of student profiles, specifying family income and savings levels, and interview the state financial aid offices directly to learn about available state grants each year.

Recent tuition increases have highlighted the importance of keeping track of the costs of college attendance, which varies in important ways by state. Total spending levels for federal and state financial aid programs tell us little about the costs for any particular student. Further, state and federal benefit formulae may interact in unknown ways. The Department of Education interviewed 50-70,000 students during the 1986-87 and 1989-90 school years to learn about multiple program eligibility. A lower cost method which would allow for more frequent observations would involve surveys of state financial aid offices, asking them to calculate student aid eligibility for various types of students. The use of profiles would also improve the reporting of the results of these large surveys.

o Report estimates of the earnings foregone by students attending college.

Policymakers have typically focused upon tuition levels and financial aid data in considering the costs of a college education. However, there is a considerable amount of cost-sharing in higher education not reflected in tuition payments. Foregone earnings totaled more than $50 billion annually (1991 dollars) since 1985, approximately 9 times the size of Pell Grant spending in 1992. Further, these costs have declined throughout the Eighties as the earnings prospects of high school graduates have dimmed.

o Experiment with questions to distinguish prior attendance at 2-year, 4-year and vocational schools as a supplement to the CPS educational attainment question.

Only a few panel datasets allow one to identify the distinct returns to different types of postsecondary education. These include the National Longitudinal Study of the High School Class of 1972, the latest follow-up of the High School and Beyond Survey and the National Longitudinal Survey of youth. While valuable, these surveys do not allow one to study potential differences in age earnings profiles beyond 14 years after high school graduation. The much larger sample provided by the CPS may prove useful in measuring these payoffs more precisely, albeit without the benefit of standardized test scores and family background as regressors.

> *o Collect UI wage record data for a targeted sample of urban youth attending urban high schools. The longitudinal surveys have typically missed a large proportion of the youth attending private vocational schools. With a more targeted sample, response rates may rise.*

While we have known little about the payoffs to a community college education until recently, we know even less about the payoffs to training at proprietary schools. Yet these schools currently claim a fifth of Pell Grant and student loan funds. The large panel surveys have not succeeded in collecting transcript information from these institutions. Rather than attempting to collect transcript and earnings data for all the schools attended by a random sample of students from a particular high school class, we may be better off with a more targeted sample of youth from a much smaller number of urban high schools and concentrate our resources to achieve higher response rates from these institutions.

Each of these recommendations is discussed in more detail below.

II. Measuring Access: How Has College Entry Varied by Family Socioeconomic Status?

The two primary sources of annual data on college enrollment are the *Current Population Survey (CPS)* and the fall enrollment estimates from the *Integrated Postsecondary Education Data System (IPEDS)*. While the CPS estimates are drawn from responses to a household survey, the IPEDS estimates are based upon an annual census of U.S. postsecondary institutions. The following discussion outlines the advantages and disadvantages of each.

Integrated Postsecondary Education Data System (IPEDS)

Since 1965, National Center for Education statistics has conducted an annual census of roughly 4000 institutions accredited at the college level (or that were not accredited but granting bachelor's degree or higher). Since 1976, in an attempt to monitor college access by race, the Office of Civil Rights has also required institutions to report biennially enrollment of students by race and gender. Figure 1 reports the IPEDS enrollment data by race.

Despite their widespread use, the IPEDS enrollment data are misleading indicators of college entry. Total student enrollment conflates at least 3 distinct factors: the duration of college enrollment, college entry and the size of the underlying population. For instance, the number of students enrolled in college increased by 14% between 1980 and 1990. However, this understated the extent of the increase in college entry since the number of youth of traditional college age, 18-24, fell by 14%. The proportion of 18-24 year-olds enrolled in college in the October CPS actually increased by 25%, almost twice as fast as the increase in total enrollment in the IPEDS.[1] Therefore, while the IPEDS data may be useful to the department of education for other reasons, it is not very informative regarding issues of college access.

The Current Population Survey

In October of each year, a supplement to the Current Population Survey questionnaire includes questions identifying the following characteristics:

- o Year of high school graduation
- o Current enrollment status at a college, university or high school
- o Grade level of current enrollment

---

[1] Bureau of the Census, P-60 Series, No. 474, page A-20.

o Type of college attended (public or private, two-year or four-year)
o Enrollment in other business or vocational coursework not part of a college or university (since 198?)

The results of the October supplement are published annually as part of the Bureau of the Census P-60 Series. Although the data are much less frequently used, the CPS also collects monthly enrollment data for all 16-24 year-old youth, indicating whether they are enrolled in high school or college and whether they are enrolled full-time or part-time.[2]

The strength of the CPS relative to IPEDS is that one observes those who are not enrolled in postsecondary institutions at the same time that one observes enrollment. Therefore, rates of enrollment of various groups--rather than simple enrollment counts--are calculable from the CPS data. For instance, one can calculate enrollment rates by age, race, region, and in many of the larger states.

However, we cannot calculate enrollment rates by family socio-economic status for students above age 19 with the current CPS. The primary obstacle is the CPS definition of household. According to the CPS Interviewer's Manual, one is considered a member of a household if one is temporarily absent from the home. Therefore, college students who periodically return to live at their family home are considered members of their parents' households. Unfortunately, one does not observe any parental information for those who move out of their parents' home but do not go to college. Therefore, family background information is missing when a youth moves out of the house permanently. Even worse, the data are missing for a non-random subset of youth: one is less likely to observe family background for the non-college-bound because they are more likely to be considered a new household. Figure 2 reports the proportion of each age group who are the children of the reference person or are other relatives of the reference person (not spouse, brother/sister, parent) by age in the 1991 Current Population Survey. Over 80% of those age 18-19 are still considered dependent members of a household. Therefore, selection bias, though still a worry, may not be hopeless for this age group. However, for those over age 19, the selection problem becomes much worse as an increasing proportion of youth have set up their own households.

Yet it would be potentially extremely valuable to keep track of gaps in college access by family economic status, given concerns about rising public tuition levels. The top two panels of Figure 3 reports the proportion of dependent 18-19 year-olds enrolled in college in each year by family income quartile and by race. The lower two panels report the difference in enrollment rates for the top three quartiles relative to the bottom quartile and by race. Gaps in enrollment rates seemed to grow by race and by income quartile throughout the Eighties.

According to the American Freshmen Survey, the average age of college freshmen has also increased over time, making it increasingly important that we be able to gauge changes in access beyond age 18-19. Given the rising importance of delayed college entry and the higher stakes in evaluating changes in college access, it may be fruitful to directly collect data on parental occupation and education on the October questionnaire, rather than continue to infer family background from the responses of other household members. For instance, questions regarding highest educational attainment and occupational status of both parents of 16-30 year-olds would provide valuable information with which to track access to college by family background even after a youth leaves the parental household.

---

[2] All age groups over 16 are asked about their primary activity during the previous week. "Attending school" is one of the possible responses. However, this category may miss those who are in school, but also working. All 16-24 year olds have been asked a separate question since 198? identifying whether they are enrolled in school, regardless of what they consider to be their primary activity during the previous week. While one cannot distinguish two-year or four-year colleges or public and private institutions with these data, the base CPS questionnaire does provide a much larger sample over the course of a year than the October questionnaire alone.

Educational Attainment vs. College Enrollment

Enrollment rates are useful for measuring the stock of students at a point in time. However, we are primarily interested in flows, i.e. knowing the proportion of a particular age group that had entered college or completed a degree by a particular age. The stock of college students will be sensitive to increases in part-time attendance and the timing of college entry. Changes in enrollment do not necessarily reflect changes in college entry. Reported educational attainment, rather than current enrollment status, provides an alternative measure of college enrollment rates.

Figure 4 reports estimates of the proportion of each cohort that reported being enrolled in college at age 18-19 and that had reported having attained some college at age 21. Though both are from the CPS data, they are based upon independent samples of the same cohorts. At least on average, the two indicators have been consistent. However, there may still be higher rates of delayed entry among more economically disadvantaged youth. Indeed, if inadequate liquidity is one reason for lower entry rates among low-income youth, we would expect more delayed entry. Kane (1991) found such evidence using the NLSY data.

However, for the same reason discussed above, CPS data will not allow one to study educational attainment beyond high school by family socioeconomic status. There is no means for measuring family background for those who are no longer dependent members of households. The Department of Education has collected longitudinal data for large samples of the high school graduating classes of 1972, 1980, 1982 and 1992. Detailed family background data allow one to study the educational attainment by family socioeconomic status. However, more than a decade has passed since the class of 1982 entered college. During that time, public tuition levels have risen more than 50%. Yet the impacts of these increases on high and low-income youth remain largely unknown. As the critical threshold for economic success moves from the high school diploma to college entry and as increasing public tuition levels rise in importance, the federal statistical mission would be well served by collecting family background data at least for young adults of college entry age.

III. College Costs: Is "Sticker Price" a Useful Indicator?

College tuition has received far more attention than college enrollment rates as an indicator of the well-being of college age youth. Each year, tuition increases at public and private universities are the subject of local and national headlines. To the extent that they accurately measure the direct costs of college, they may measure one of the barriers to college education. In this section, I evaluate to what extent these "sticker prices" can be taken as an accurate measure of the costs of college enrollment.

Institutional Aid

In addition to collecting data on tuition and required fees, the Integrated Postsecondary Education Data System asks colleges and universities to report expenditures on institutional aid to college students. Figure 5 reports for public and private 4-year institutions average annual tuition and required fees and institutional financial aid per full-time-equivalent (fte) student.

Between 1980 and 1990, average public 4-year tuition and required fees increased by 50% after accounting for inflation. Over the same period, institutional aid at public 4-year institutions also increased by 50%. Private institutions have been raising institutional aid expenditures more rapidly than they have been raising tuition levels.

Our statistical system evolved at a time when there was little means-tested aid. Published tuition and required fees represented a good indicator of the direct costs of college enrollment. Federal means-tested aid--the bulk of means-tested aid--was more easily observed directly using the program rules. As a result, the data reporting system does not distinguish among different types of institutional aid awarded. For instance, the IPEDS estimates in Figure 5 include graduate student aid as well as aid to undergraduates. It also includes merit-based as well as means-tested aid. Therefore, it is not possible with the IPEDS data to identify means-tested institutional aid to undergraduates. Indeed, doing so may impose significant burdens on college administrators whose time is already taxed by the reporting demands of IPEDS.

State and Federal Means-tested Financial Aid

However, institutional grant aid is not the only source of measurement error in using tuition to signal the direct costs of college. The Washington Office of the College Board has provided a useful public service in collecting and publishing data annually on the amount of spending on various state and federal programs over time. Figure 6 reports these figures for each year since 1970. Guaranteed student loans are the largest spending category. Although some proportion of these loans are subsidized in two ways--with below-market interest rates and deferred repayment--the implicit subsidy value of the loan programs is not typically reported. In fact, it is often difficult to compute, given the range of interest rates available in different programs.

Pell Grants (Basic Educational Opportunity Program) represent the largest means-tested grant program. As reported in Figure 6, real spending on Pell Grants grew during the Eighties. However, these figures give a misleading impression of the amount of aid available to the neediest students. Changes in program rules have expanded the eligibility of middle income students and take-up rates have increased as enrollment rates grew. A more useful measure of the amount of Pell Grant aid available to the neediest youth is the Pell Grant maximum award, which declined by 15% in real value between 1980 and 1990, as public tuition levels increased.

To supplement federal spending, a number of states have need-based grant programs. Figure 6 reported that these programs totaled roughly $2.1 billion in 1991-92, one-third the size of Pell Grants. Unfortunately, total need-based state grant spending does not tell us much about the costs of college enrollment for many of the same reasons that total Pell Grant program spending reveals little. Further, there is no single measure such as the maximum Pell award with which to summarize the costs of college enrollment. Aid programs differ dramatically across states. For instance, New York, Minnesota and Vermont spent $294, $192 and $177 per 18-24 year old youth in means-tested state aid in 1992. In contrast, twenty-two states spent less than 25 dollars per 18-24 year-old youth in that year.[3] We need to learn more about the distribution of state grant aid and their interaction with federal aid programs.

National Postsecondary Student Aid Surveys

State, federal and institutional financial aid programs overlap in ways that are often not obvious even to the policy-makers involved. Benefit formulas and amounts vary across programs. Students may receive Pell Grant aid, state grant aid and an institutional grant as well. Totalling the tuition amounts as well as grants received under various programs tells us only the average direct cost of a college education and reveals nothing about the distribution of costs. Yet it is the distribution of direct costs to students, particularly to those at the low end of the family income distribution, that concerns us.

To fill this gap, the Department of Education collected data from roughly 60,000 college students during the 1986-87 academic year and 70,000 students in 1989-90.[4] The resulting reports contain estimates of the costs and types of aid received by a number of categories: public and private universities, 2-year and 4-year colleges, full-time and part-time students, race, family income level for dependent students. However, the published data are often difficult to interpret. For instance, it was not possible in the published tabulations to calculate the net tuition costs of students attending public and private 4 year colleges, after accounting for grant aid.

---

[3] These figures were drawn from Davis (1993).

[4] For a detailed description of the results of these surveys, see the National Center for Education Statistics (1988), National Center for Education Statistics (1993a), National Center for Education Statistics (1993b).

An Alternative Method of Reporting and Collecting Net College Cost Data

An alternative method of reporting the NPSAS data would involve the use of "student profiles." One could construct 3 different profiles: one for a representative disadvantaged youth, a "middle income" youth and a "high income youth". In specifying the profiles, one would want to specify and hold constant each of the characteristics relevant to state and federal financial aid. For instance, the disadvantaged youth may be attending a public 2-year institution full-time, have poverty level family income, 2 siblings not in college, \$1,000 in personal savings, \$2,000 in parental savings and \$0 in housing wealth. The "middle income" youth may be chosen to represent the median of each of these characteristics, the "high income" youth the 75th percentile. The mean direct costs faced by students with these profiles should then be reported by region or state (there is no such reporting in the summary reports at this point) and over time. (If the sample sizes are small, one could use income and family size ranges.) The dollar figures, such as family income, would of course be indexed, but would not change with the distribution of family income of college students. The advantage of this approach is that it allows us to know precisely what is being held constant in making these comparisons.

A second advantage is that it would provide for a low-cost alternative in collecting data on state aid in the years between the national surveys. We may not need to interview 60-70,000 students to learn about how different financial aid programs overlap. The National Association of State Scholarship and Grant Programs surveys its members annually, collecting data on total spending levels in various grant programs. However, as mentioned above, the spending levels themselves need not reflect the benefits available to youth of any particular income level. Among other things, they would simply reflect differences in the income distributions across states. Therefore, the state financial aid authorities could be asked to calculate the aid available for each of the representative youths. With data on tuition costs and knowing the Pell Grant rules, it would be a straightforward calculation to keep track of the direct costs of postsecondary education during the interim years as well as provide a check on the data collected in the NPSAS surveys. And it could be done without interviewing 60-70,000 students each year. The only portion of direct costs that would be difficult to observe would be institutional aid, which tends to be low at public institutions. If public tuition levels are the price relevant to the marginal student, even this may not be an important limitation.

The NPSAS will continue to be valuable to the extent that some state aid programs are discretionary, campus-based programs, which, like institutional aid, would be missed in such a survey. Further, the NPSAS tells us about take-up rates. Though the data are not yet available, the survey was conducted again in 1991-92, with plans for another in 1995-96.

Foregone Earnings

In thinking about the costs of postsecondary education, higher education analysts have focused upon the components of the direct costs: tuition levels, direct subsidies by state and local governments, and financial aid. The statistics reported in the federal statistical digests reflect this perspective. However, one component of the costs of postsecondary education which is often ignored are the earnings foregone by the students themselves. Although they do not appear on any public budgets, such opportunity costs represent a large share of the costs of postsecondary education and they have been changing over time.

To estimate the size of these indirect expenditures, I estimated the following regression equation using earnings data for 18-24 year-olds from the outgoing rotation groups for each month of the CPS:

$$y_i = X\beta + \delta_{ft} \mathit{Enrolled\ in\ College\ FT}_i + \delta_{pt} \mathit{Enrolled\ in\ College\ PT}_i + \epsilon_i$$

where X_i includes age dummies, race and gender dummies and dummies for number of years of school completed. The dependent variable was reported average weekly earnings, including those with zero incomes. The sample consisted of high school graduates who were either enrolled in college or in the labor force. The coefficients on school enrollment, δ_{ft} and δ_{pt}, are rough estimates of the earnings foregone by those attending college full-time and part-time respectively. Further, they are likely to be lower bounds, since one might have expected college students to have had higher earnings than non-college-bound youth if they had not been enrolled themselves, given that the average college student has better high school performance, pre-college ability test scores and more favorable family backgrounds than the non-college bound.

Therefore, one could estimate the costs of college enrollment by multiplying δ_{ft} and δ_{pt} by the number of youth enrolled in college part-time and full-time. Table 1 contains the resulting estimates of opportunity costs and the earnings foregone per student. The earnings foregone by college students in a given year is estimated to total over 50 billion dollars (1991 dollars) in each year since 1985. Foregone earnings are roughly 9 times the size of total Pell Grant spending in 1992 and considerably higher than the value of the state subsidies paid to postsecondary education. There is a considerable amount of cost-sharing by students in postsecondary education even beyond the tuition they pay.

IV. Measuring the Payoffs to College

The Current Population Survey has served well as an indicator of changes in the payoff to education. While the college-high school earnings differential grew after 1979, a number of published papers had noted the fact by the late Eighties.[5] Data published by the Bureau of the Census and the Bureau of Labor Statistics provide an opportunity to monitor earnings differences between college graduates and high school graduates. For instance, the January issue of Employment and Earnings and the annual Census publication, Money Income of Households, Families and Persons in the United States, provide readily available sources.

Identifying the Payoffs to Different Types of Postsecondary Education

However, our statistical system has performed much less well in identifying the returns to alternative types of postsecondary education. For instance, although they currently enroll over half of first-time freshmen, we have known very little about the economic payoffs to a community college education until recently. This has been a particularly regrettable gap in the literature, because community colleges enroll a disproportionate share of those students whose enrollment decisions are affected by state and federal financial aid policies.[6]

One important obstacle has been the lack of retrospective data on the type of postsecondary institution attended. While the October CPS asks *current* students to identify where whether their schools are two-year or four-year colleges or vocational schools, one cannot distinguish between non-students who have attended a two-year or four-year institution. For 50 years between the 1940 and 1990 decennial census, the standard Bureau of the Census educational attainment question merely asked for the highest grade attended and whether the respondent completed that grade. A CPS supplement which asked respondents to identify the type of college attended should be considered. Such a question would require experimentation, however, because any particular student may have attended varying amounts of the each type of school.

An alternative approach is to continue to rely upon the postsecondary transcript data collected for the high school classes of 1972 and 1982. (The latter will be available in late 1994.) Recently, Kane and Rouse (1993), Grubb (1993) and Hollenbeck (1992) have studied the payoffs to alternative types of postsecondary education using postsecondary transcripts provided by the National Longitudinal Study of the High School Class of 1972. Before controlling for family background and measured high school performance, Kane and Rouse

---

[5] For a review of this literature, see Levy and Murnane (1992). They cite a number of papers on the leading edge of what became a cottage industry. (Levy (1988), Murphy and Welch (1989)).

[6] Even though they account for roughly a quarter of Pell grant spending and a fifth of guaranteed loan volume, these figures are certainly underestimates of the proportion of students who would not have gone to college in the absence of aid. Presumably, a higher proportion of four-year college age recipients would have attended some college in the absence of aid. In simulations reported by Manski and Wise (1983), community college entrants accounted for two-thirds of those who would not have entered college in the absence of the Pell Grant program.

(1993) find that a full year of college credits (using 30 semester credits per year as a rule of thumb) increases male and female earnings by roughly 6 and 7-8 percentage points respectively. After including family background, high school class rank and standardized test scores, these estimates fall by 13-20%. Each year of schooling was associated with a 4-7% earnings differential 14 years after high school. As mentioned above, the strength of the NLS-72 data is that it allows one to observe the distinct payoffs to two-year and four-year college credits. For those not completing college, Kane and Rouse estimate that a year at a two-year or four-year college would lead to a 6% and 3% increase in annual earnings for men respectively and a 7% and 10% increase in annual earnings for women. These differences in the payoffs to two-year and four-year college credits are not statistically significant. After correcting for computational errors in Grubb (1993),[7] one finds estimates of similar magnitude as reported by Kane and Rouse.

The high school class of 1972, upon whose experience most of the above evidence depends, graduated from high school more than two decades ago. Community colleges today are quite different. Certainly, among those completing associate degrees, there was a shift toward toward vocational subjects between 1972 and 1982, although the distribution seems to have stabilized after 1982.[8] Kane and Rouse (1993) also report results from the National Longitudinal Survey of Youth, with youth who were 14-21 in 1979. After controlling for family background, Armed Forces Qualifying Test scores and work experience, even those who had attended community colleges without finishing had earnings 4-13% more than similar high school graduates.[9]

The estimates from the 10-year follow-up of the high school class of 1982 will shed more light when they become available later this year. Such long-term panel data have several advantages: they measure pre-college differences in test scores and family background as potential regressors; they observe type of school contemporaneously with attendance and collect transcripts to lessen errors from self-reporting;[10] they observe long-term earnings differences 10 to 14 years after high school. However, an important disadvantage is that they are costly data to collect. After these data from the high school class of 1982, we may not have another chance to observe the payoffs to different types of postsecondary education for another 10 years, until the National Educational Longitudinal Survey of the class of 1992 conducts a long-term follow-up survey.

---

[7] For a detailed description of these differences, see Kane and Rouse (1994).

[8] Grubb (1994) emphasizes the differences in labor market payoffs to "vocational" and "academic" credits completed in postsecondary institutions. In particular, he reports higher estimated payoffs to vocational credits for men and academic credits for women. However, despite the differences in these point estimates, tests of statistical significance would not lead one to reject the hypothesis that the payoffs were the same at conventional levels (although one could reject the hypothesis of similarity at the .10 level for men). As Grubb has pointed out, one's interpretation of these facts depends upon the null hypothesis being tested. There is simply too little evidence to settle the matter.

[9] Grubb (1994) uses the Survey of Income and Program Participation to evaluate the returns to schooling. Although there is no family background or ability measure, the survey allows one to identify wage differentials by year of schooling completed and degrees received. Males with less than 1 year of college and no credential earned 4-12 percentage points more than high school graduates in 1984 and 1987. (The estimates in the lower end
of this range are not statistically significant.) Males with 1 year of college earned 12-16% more than high school graduates. Females with 1 year of college earned 10% more than high school graduates, although the estimate is only marginally statistically significant.

[10] Although transcript measurement error may be just as important as self-reporting error, as reported in Kane and Rouse (1993).

Identifying the Payoffs to Education for Drop-outs and Short-term Students

Particularly because only a quarter of two-year college entrants from the class of 1972 completed an associate's degree and less than 40% went on to 4-year colleges, questions regarding the value of degree completion beyond the number of credits completed are important. Indeed, approximately 40% of those who dropped out of community colleges did so after less than a semester in credits.

Beginning with the 1990 Decennial Census and the 1992 Current Population Survey, the Bureau of the Census changed the coding of educational attainment to reflect degree completion, rather than the number of years of schooling completed. This may seem to have been a step forward. But the Bureau of the Census simultaneously took an equally large step back. Unfortunately, the number of years of schooling completed for those not completing a college degree was lost in the transition. All those with "some college, no degree" were lumped together in one category.

However, when the new question was tested in the February 1990 Current Population Survey, it was possible to combine information on years of school completed with the degree attainment questions and ask whether those with two years of college and no degree earned any less than those with the same amount of schooling and an associate degree. Table 2 contains the tabulations of these data reported by Paul Siegel (1991) of the Bureau of the Census.[11] Inspecting the row corresponding to those with 2 years of college, it is evident that there were essentially no differences in earnings among those with no degree or an associate's degree. All three groups earned roughly $28,000 per year in 1990. More recent work by Jaeger and Page (1994), using matched CPS data from 1991 (including the question regarding years of schooling completed) and 1992 (with data on degree completion) suggests larger AA degree effects for women and BA effects for men. There was also evidence of a BA degree effect for women.

The Bureau of Labor Statistics is currently experimenting with further revisions to the educational attainment data which would recapture the information on the number of years of schooling completed by those without degrees. Given the importance of observing the payoffs for short-term non-completers, the new data will be eagerly awaited. However, these will not be available until 1996 at the earliest.

Value-Added Measures

Unfortunately, the CPS does not allow one to adjust for differences in family background and ability between degree completers, college drop-outs and other high school graduates. Further, even the results from the NLS-72 and the High School and Beyond depend upon the use of measures of family background, high school performance and standardized test scores as regressors to control for any unobserved differences between 2-year or 4-year college students and high school graduates. These measures may only imperfectly capture the differences between high school graduates and college entrants. For instance, one common worry is that college graduates are simply more "motivated" and may have had higher earnings even without entering college.

In other fields of social policy, experimental evaluation with random assignment has been the answer to the empirical quandary. Indeed, there have been two recent experimental evaluations of classroom training for youth. A subset of the sample in the Job Training Partnership Act evaluation received classroom training. The JOBSTART evaluation provides another example. (For a summary of these findings, see Orr et al. (1994) and Cave et al. (1993)).

Though these results have often been interpreted as showing a low payoffs to classroom training for disadvantaged youth, there is absolutely no basis in the data for such an inference. The "treatment" being evaluated is not classroom training per se. In the case of JTPA, the treatment was simply the addition of JTPA services to the current menu of options, which included attending the same community colleges used by the JTPA program, but funded by the Pell Grant program. As reported in Kane (1994), the treatment group at the

---

[11] Hungerford and Solon (1987) also report finding non-linearities in the payoff to schooling at 8, 12 and 16 years of schooling.

15 JTPA sites evaluated would have expected to pay -$277 to attend their local community college and $60 to attend the 4-year public colleges in their states. Therefore, the direct costs of college entry were quite low for these youth in the absence of JTPA, given the availability of Pell Grants. This may explain the 50% enrollment rate of the control group. Even if there were a huge payoff to postsecondary education for some subset of eligible youth, we would have observed small effects of the "treatment" in these recent experiments.

In contrast, these studies often find payoffs to providing job search assistance. Rather than reflecting the relatively high payoffs to job search, such evidence may simply be do the lack of availability of such services elsewhere.

Experimental evaluations will only work in postsecondary education if we are willing to deny access to postsecondary education to the control groups. This has not been done to date. However, the empirical challenges differ from those faced by welfare employment and training program evaluators. First, there are many more youth already participating in classroom training than there were welfare recipients in job search assistance. The costs of denying access are higher. Second, currently high enrollment rates among youth provide access to alternative forms of evidence, such as provided by the panel data discussed above. The value of the information gained through experiments is almost certainly lower, because there are alternative estimates. Third, there is exogenous variation in access to college due to state tuition differences and the distance of one's high school from the closest college. Both of these sources of variation could be exploited further (current examples are Kane and Rouse (1993) and Card (1994)), in addition to controlling for observed differences in family background and student test scores.

Other Approaches

Recent work by Jacobson, LaLonde and Sullivan (1994) evaluating the payoffs to postsecondary training for displaced workers provides another model. Rather than following a sample of high school graduates over time and eliciting data from all the postsecondary schools they attended, the authors collected the unemployment insurance wage records for a sample of displaced workers attending a particular community college in Pittsburgh as well as a sample of other displaced workers in the same area who did not.

One advantage of such an approach is that the task of tracking down transcripts from a number of different schools is lessened. Second, the strategy relies upon observing wages before and after college entrance, to allow the investigators to control for fixed person effects. Although this strategy may work for later college entrants, it does not serve well for students immediately after college since they have no wage history. Nevertheless, this approach may be quite helpful for studying private vocational schools, which have not been adequately evaluated with other data. This issue is discussed at more length in the next section.

V. Private Vocational School Students

We are beginning to obtain better information on the characteristics of students enrolled in private, for-profit vocational schools or proprietary schools. Beginning in 1986, the IPEDS universe was expanded to include 5,694 private, for-profit, less than 2-year institutions. Indeed, the Higher Education Reauthorization Act of 1992 required all institutions receiving financial aid to respond to the IPEDS survey, although that provision has not been enforced. By 1990, response rates of these institutions to the IPEDS enrollment queries was quite high, 92.6%. Response rates have been much lower on the financial portion of the questionnaire, 65-70%.[12]

Using the 1987 IPEDS list of institutions as its universe, the 1990 NPSAS sample contained 8,065 students attending proprietary schools. Given that the response rates were quite high among these institutions

---

[12] These response rates were obtained during a phone conversation with Vance Grant of the National Center for Education Statistics.

(87%) (NCES, (1992)), we have a chance to learn something about the characteristics and methods of financing used by these students.[13]

Beginning in the late Eighties, the October Current Population Survey also distinguished between proprietary school students and those attending 2-year and 4-year colleges. The CPS question in 1992 read as follows:

> "Excluding regular college courses and on-the-job training, is .... taking any business, vocational, technical, secretarial, trade or correspondence courses?"

However, the question captured a different population than the IPEDS estimates. In Fall of 1992, the IPEDS data suggested that there were approximately 800,000 students enrolled in less-than-two-year institutions. In October 1992, 3.4 million respondents reported that they were taking vocational courses but were not enrolled in college.[14] It is not clear that this implies that the IPEDS data are grossly understated or if the CPS question is capturing those enrolled in courses offered by institutions other than proprietary schools. These potential inconsistencies deserve to be resolved by the Bureau of the Census and the Department of Education.

However, the situation is much worse for evaluating the payoffs to proprietary schools. The Department of Education monitors loan default rates, which are higher at proprietary schools than other types of postsecondary institutions. These have often been interpreted as suggesting that the education being provided at proprietary schools is not worth the tuition costs born by students. However, the higher default rates may simply be due to the socioeconomic backgrounds of the students attending the schools. This is precisely the same issue as so often is raised in the evaluation of K-12 schools: are low test scores measuring low value-added or are they measuring the low starting point of their students?

Unfortunately, the panel data-sets are not very useful in examining the payoff to a proprietary school education simply because the response rates from proprietary schools which have been asked to supply transcript information are quite low. In the NLS-72 transcript survey, the response rates for proprietary schools was 43%. Two-year and four-year colleges had response rates over 90%.[15] One reason for the difficulty is that the schools often close. Roughly a quarter of the non-response was confirmed to be due to school closings. Another problem is poor record-keeping. Half of the non-reporting schools reported that the records were lost or destroyed. The transcript collection effort for the High School and Beyond sophomore cohort is ongoing. Despite efforts to keep response rates high, proprietary schools are projected to have response rates below two-thirds.

Rather than attempting to identify a nationally representative sample of high school students and then hunting down transcripts from the thousands of postsecondary institutions these students enter, an alternative strategy is to oversample high school students attending schools which are most likely to produce prospective proprietary school students and follow them. To the extent that many entrants may have sporadic employment histories before entry, the collection of high school grades or test scores would provide a more adequate basis for estimating value-added measures. Although the estimates may be less easily generalizable to proprietary schools outside these areas, the result is likely to teach us more about the payoffs to proprietary schools than the current system with very low response rates.

---

[13] For a more detailed description, see NCES (1993a).

[14] Figures were drawn from Bureau of the Census, School Enrollment: Social and Economic Characteristics of Students, October 1992 P-20 Series, No. 474, p. 47.

[15] Jones, et al. (1986)

VI. Summary and Conclusions

Increasingly, college entry, rather than high school graduation, has become the focus of public concern, as the labor market has changed. Our statistical collection efforts should adapt to this change. There are four primary weaknesses in the current system:

> *Beyond age 19, we are can not reliably observe differences in school and enrollment by family background. Socioeconomic differences in college entry rates and attainment should be monitored annually or biennially.*
>
> *While there is plentiful data on average college costs because we can observe total spending on various types of financial aid, we have traditionally known little about the distribution of net costs, particularly for low-income youth. Recent surveys of undergraduates have begun to fill the gap. However, they have often not been reported in a useful way. The paper recommends the use of student profiles for collecting such data more frequently from state financial aid agencies.*
>
> *Foregone earnings are typically ignored in the calculation of the costs of postsecondary education. Yet this is the primary method by which youth share the costs of postsecondary education.*
>
> *We have poor data on the payoffs to alternative types of postsecondary education. Better data could be pursued with a CPS supplement asking respondents to report the type of schools attended--2-year, 4-year or proprietary school--rather than just the number of years of school completed.*
>
> *Proprietary schools receive considerable public resources but have received little public scrutiny. The traditional panel data sets (NLS-72, HSB and NLSY) are not yielded considerable evidence. Given their importance, more targeted efforts should be directed at observing the payoffs at these institutions. Social experiments are not likely to help.*

Our statistical system evolved when tuition was an adequate measure of the direct college costs of college and when the differences among institutions were not as large. Now that the stakes have been raised by the rising value of a college education, the investment in upgrading the data we collect is necessary.

Table 1
Estimating the Value of Earnings Foregone by College Students
(1991 Dollars)

Year:	Monthly Earnings Losses During College Enrollment: (Std Error)		Avg. Number of 18-24 Year-Old Youth Enrolled in College per Month: (millions)		Estimated Foregone Earnings (billions)
	Part-time	Full-time	Part-time	Full-time	
85	-150 (16.7)	-815 (9.7)	.9	5.1	51.5
86	-161 (18.1)	-833 (10.2)	.9	5.0	51.7
87	-174 (17.5)	-862 (10.3)	1.0	5.1	54.8
88	-184 (17.3)	-842 (10.6)	1.0	5.1	53.7
89	-233 (16.7)	-895 (8.8)	.9	5.3	59.4
90	-253 (15.5)	-878 (8.4)	1.0	5.3	58.9
91	-281 (15.5)	-823 (8.4)	1.0	5.3	55.7

Note: All of the above were estimated using the sample of 18-24 year-old high school graduates in the outgoing rotation groups of the CPS, 1985-91. Average weekly earnings were converted into monthly earnings by multiplying by 4.33. Only those who were in school or were in the labor force were used in the sample. The separate specification for each year included gender, race, single year of age dummies and dummies for years of education attended.

Table 2

Mean Earnings by Years of College Completed and Degrees Received in the February, 1990 CPS

(Standard Errors are in parentheses)

Reported Degree Attainment

Years of College Completed	No College	Some College, No Degree	Occupational Associate's	Academic Associate's	Bachelor's	Master's
6+					35252 (1019)	42091 (553)
5					35276 (835)	40145 (1301)
4		32245 (1771)	31135 (1674)	34385 (1998)	35992 (312)	37709 (2146)
3		30224 (771)	29735 (1025)	30845 (1424)	30923 (1930)	
2		28928 (413)	28139 (535)	28558 (603)		
1		26057 (348)	24199 (1571)			
High School Graduate	23858 (135)	25802 (480)	23310 (1374)			

After including gender, race, and age as regressors:

Years of College Completed	No College	Some College, No Degree	Occupational Associate's	Academic Associate's	Bachelor's	Master's
6+					32608 (837)	38772 (439)
5					33102 (758)	37909 (901)
4		31100 (1350)	30707 (1700)	32771 (1480)	34124 (290)	35036 (1592)
3		27738 (634)	29184 (1322)	29813 (1377)	29791 (1954)	
2		27264 (413)	26534 (660)	27245 (674)		
1		24056 (397)	23988 (2042)			
High School Graduate	22070 (223)	24382 (516)	23339 (2116)			

Hypothesis Tests:
(p-values)

Years Zero within MA	.060	MA=BA within Years	.001
Years Zero within BA	.030	BA=Acad AA within Years	.660
Years Zero within Acad AA	.001	Ac AA=Occ AA within Years	.670
Years Zero within Occ AA	.010	Occ AA-No Deg within Years	.810
Years Zero within Some Coll	.000		

Note: Data are drawn from Paul M. Siegel, "Note on the Proposed Change in the Measurement of Educational Attainment in the Current Population Survey" U.S. Bureau of the Census, Draft, February 5, 1991.

References

Ashenfelter, Orley and David Card, "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs" *Review of Economics and Statistics* (1985) Vol. 67, No. 4, pp. 648-660.

Ashenfelter, Orley and Alan B. Krueger "Estimates of the Economic Return to Schooling for a New Sample of Twins" Princeton University Industrial Relations Section Working Paper No. 304, July 1992.

Barnow, Burt S. "The Impact of CETA Programs on Earnings: A Review of the Literature" *Journal of Human Resources* (1987) Vol. 22, No. 2, pp. 157-193.

Bloom, Howard S., Larry L. Orr, George Cave, Stephen H. Bell, Fred Doolittle and Winston Lin *National JTPA Study Overview: Impacts, Benefits and Costs of Title II-A* A Report to the U.S. Department of Labor, January, 1994.

Card, David "Using Geographic Variation in College Proximity to Estimate the Return to Schooling" NBER Working Paper No. 4483, October 1993.

Cave, George, Hans Bos, Fred Doolittle and Cyril Toussaint Jobstart: Final Report on a Program for School Dropouts (New York: Manpower Demonstration Research Corporation, 1993).

College Board. Trends in Student Aid: 1984 to 1994 (New York: The College Board, 1994).

College Board. Trends in Student Aid: 1963 to 1983 (New York: The College Board, 1983).

Davis, Jerry. National Association of State Scholarship and Grant Programs, Annual Survey Report: 1992-93 Academic Year (Harrisburg, PA: Pennsylvania Higher Education Assistance Agency, 1993).

Griliches, Zvi "Estimating the Returns to Schooling: Some Econometric Problems" *Econometrica* (1977) Vol. 45, pp. 1-22.

Grubb, W. Norton "The Varied Economic Returns to Postsecondary Education: New Evidence from the Class of 1972" *Journal of Human Resources* (1993) Vol. 28, No. 3, pp. 365-382.

Grubb, W. Norton "Postsecondary Education and the Sub-Baccalaureate Labor Market: Corrections and Extensions" Unpublished Paper, February 1994.

Grubb, W. Norton "The Returns to Education and Training in the Sub-Baccalaureate Labor Market: Evidence from the Survey of Income and Program Participation" Unpublished Draft, National Center for Research in Vocational Education, University of California at Berkeley, March, 1994.

Hollenbeck, Kevin "Postsecondary Education as Triage: The Consequences of Postsecondary Education Tracks on Wages, Earnings and Wage Growth" Paper presented at the Western Economic Association Meetings, San Francisco, 1992.

Hungerford, Thomas and Gary Solon "Sheepskin Effects in the Returns to Education" *Review of Economics and Statistics* (1987) Vol. , pp. 175-177.

Jaeger and Page, "Degrees Matter" Unpublished Paper, Department of Economics, University of Michigan, 1994.

Jones, Calvin, Reginald Baker, Robert Borchers et al. National Longitudinal Study of the High School Class of 1972, Postsecondary Education Transcript Study, Data File User's Manual (Chicago: National Opinion Research Center, 1986).

Kane, Thomas J. "College Entry by Blacks Since 1970: The Role of College Costs, Family Background and the Returns to Education" PhD dissertation, Harvard University, 1991.

Kane, Thomas J. and Cecelia Rouse "Labor Market Returns to Two- and Four-Year Colleges: Is a Credit a Credit and Do Degrees Matter?" Kennedy School of Government Faculty Research Working Paper Series R93-38 February, 1993. (Forthcoming *American Economic Review*, 1995)

Kane, Thomas J. and Cecelia Rouse "Comment on W. Norton Grubb" Forthcoming *Journal of Human Resources*, 1994.

Kane, Thomas J. "College Entry by Blacks Since 1970: The Role of College Costs, Family Background and The Returns to Education" Kennedy School of Government, Faculty Research Working Paper Series, No. R94-7, February, 1994. (Forthcoming *Journal of Political Economy*, 1994)

Kane, Thomas J. "Recognizing Experimental and Non-experimental Estimates of the Returns to Schooling" mimeo, Brookings Institution, 1994.

Katz, Lawrence and Kevin Murphy "Changes in Relative Wages, 1963-1987: Supply and Demand Factors" NBER Working Paper 3927, December 1991.

Levy, Frank. "Incomes, Families and Living Standards" in American Living Standards Threats and Challenges Robert Litan, Robert Lawrence and Charles Schultze (eds.) (Washington: Brookings Institution, 1988).

Levy, Frank and Richard Murnane "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations" *Journal of Economic Literature* (1992) Vol. 30, No. 3, pp. 1333-1381.

Manski, Charles F. and David A. Wise *College Choice in America* (Cambridge: Harvard University Press, 1983.)

Murphy, Kevin and Finis Welch. "Wage Premiums for College Graduates: Recent Growth and Possible Explanations" Educational Researcher (May, 1989) pp. 17-26.

Murphy, Kevin and Finis Welch "Industrial Change and the Rising Importance of Skill" in Sheldon Danziger and Peter Gottschalk (eds.) *Uneven Tides* (New York: Russell Sage Foundation, 1993).

National Center for Education Statistics, Undergraduate Financing of Postsecondary Education: A Report of the 1987 National Postsecondary Student Aid Study (Washington: Office of Educational Research and Improvement June, 1988).

National Center for Education Statistics, Methodology Report for the 1990 National Postsecondary Student Aid Survey (Washington: Office of Educational Research and Improvement, May, 1992).

National Center for Education Statistics, Profile of Undergraduates in U.S. Postsecondary Institutions: 1989-90 (Washington: Office of Educational Research and Improvement, February, 1993a).

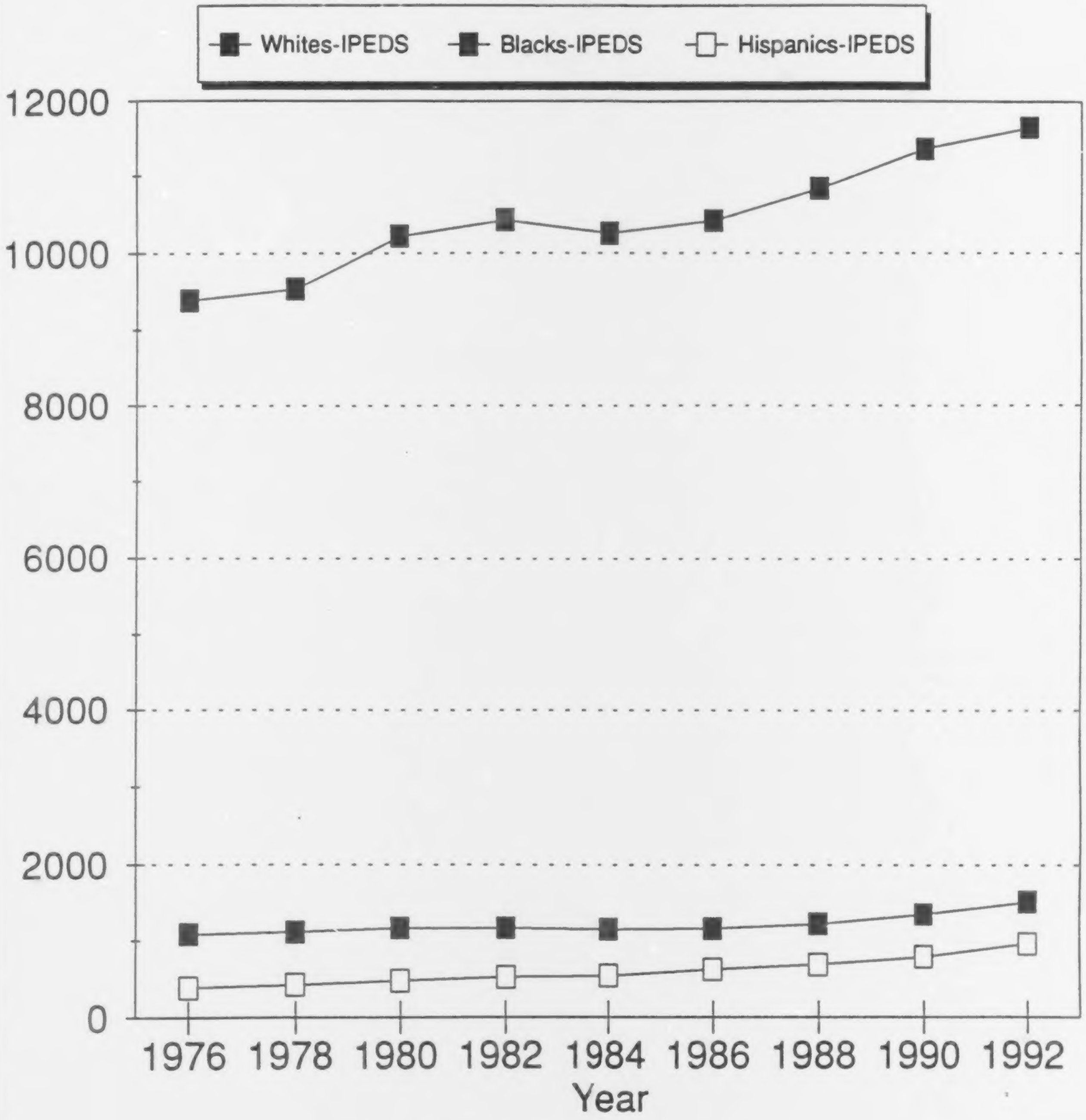
National Center for Education Statistics, Financing Undergraduate Education: 1990 (Washington: Office of Educational Research and Improvement, May, 1993b).

National Center for Education Statistics, Changes in Undergraduate Student Financial Aid: Fall, 1986 to Fall, 1989 (Washington: Office of Educational Research and Improvement, August, 1993c).

Orr, Larry L., Howard S. Bloom, Stephen H. Bell, Winston Lin, George Cave and Fred Doolittle *National JTPA Study: Impacts, Benefits and Costs of Title II-A* Draft report to the U.S. Department of Labor, March, 1994.

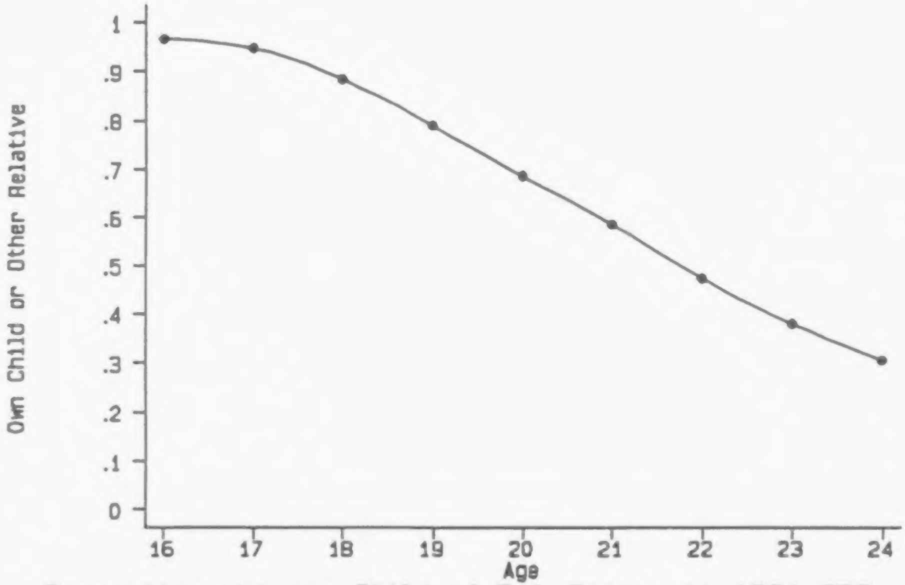
Siegel, Paul M. "Note on the Proposed Change in the Measurement of Educational Attainment in the Current Population Survey" U.S. Bureau of the Census, Draft, February 5, 1991.

Figure 1: Enrollment by Race-IPEDS



Source: "Digest of Education Statistics," Table 203 from 1994 and Table 174 from 1990, National Center for Education Statistics. 81% of Hispanics assumed to be white, 12% of Hispanics assumed to be black.

Figure 2.



Proportion who are Child of Ref. Person in 1991 CPS

Figure 3.

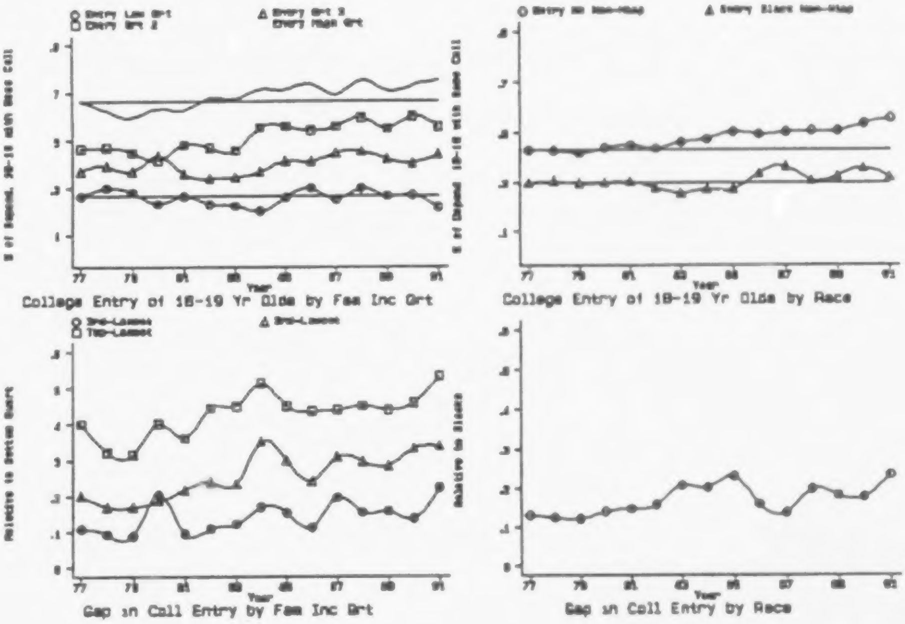
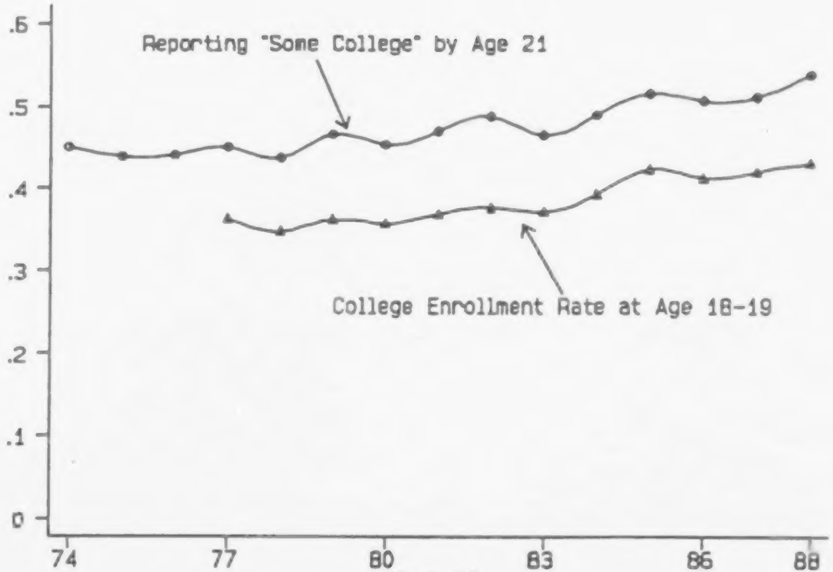
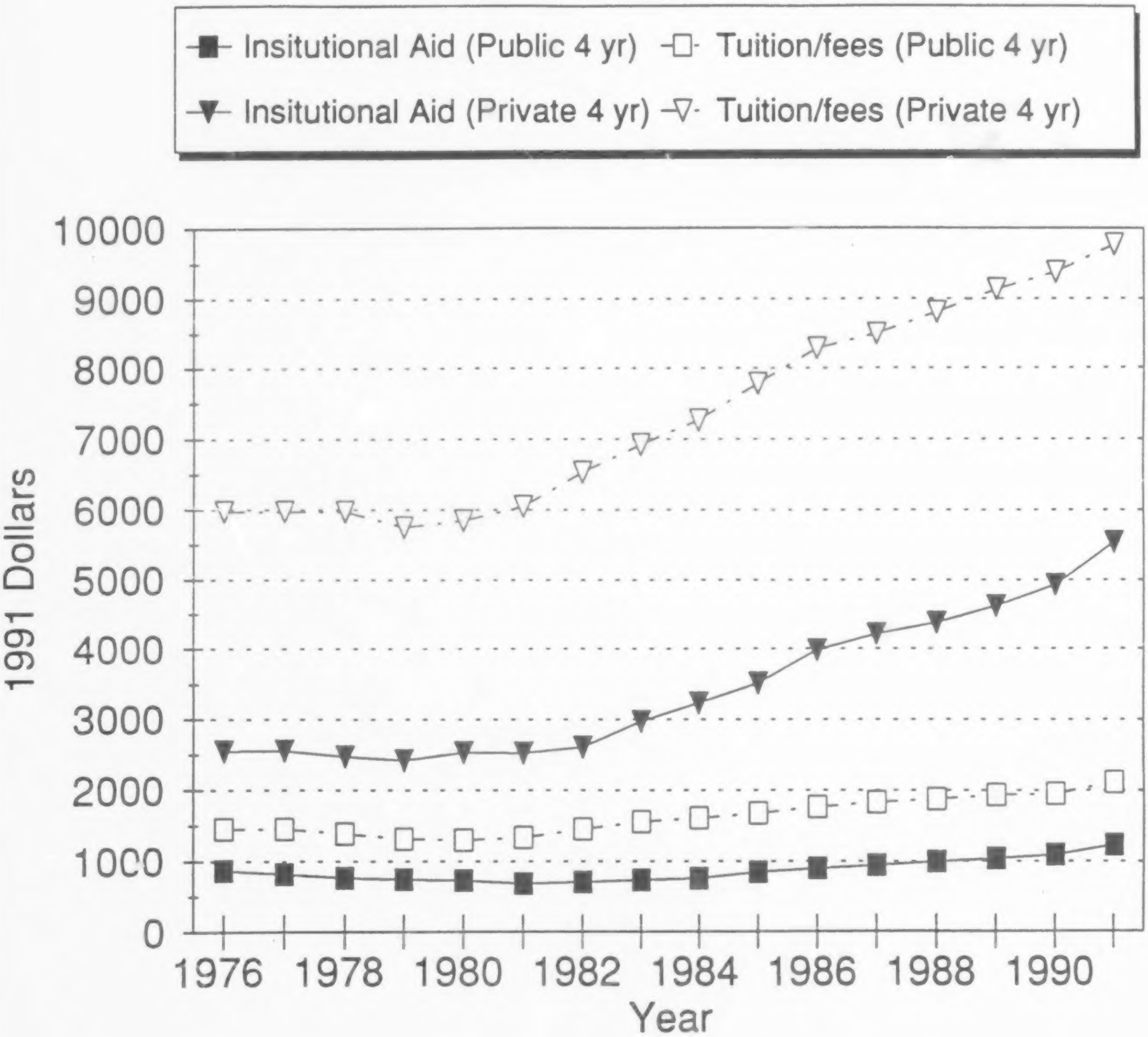

Entry Low Qrt
Entry Qrt 2
Entry Qrt 3
Entry High Qrt
% of Depend. 18-19 with Some Coll
.8
.7
.5
.3
.1
77
79
81
83
85
87
89
91
Year
College Entry of 18-19 Yr Olds by Fam Inc Qrt
Entry Wh Non-Hisp
Entry Black Non-Hisp
College Entry of 18-19 Yr Olds by Race
2nd-Lowest
Top-Lowest
3rd-Lowest
Relative to Bottom Quart
.6
.5
.4
.3
.2
.1
0
Gap in Coll Entry by Fam Inc Qrt
Relative to Blacks
Gap in Coll Entry by Race

Figure 4.



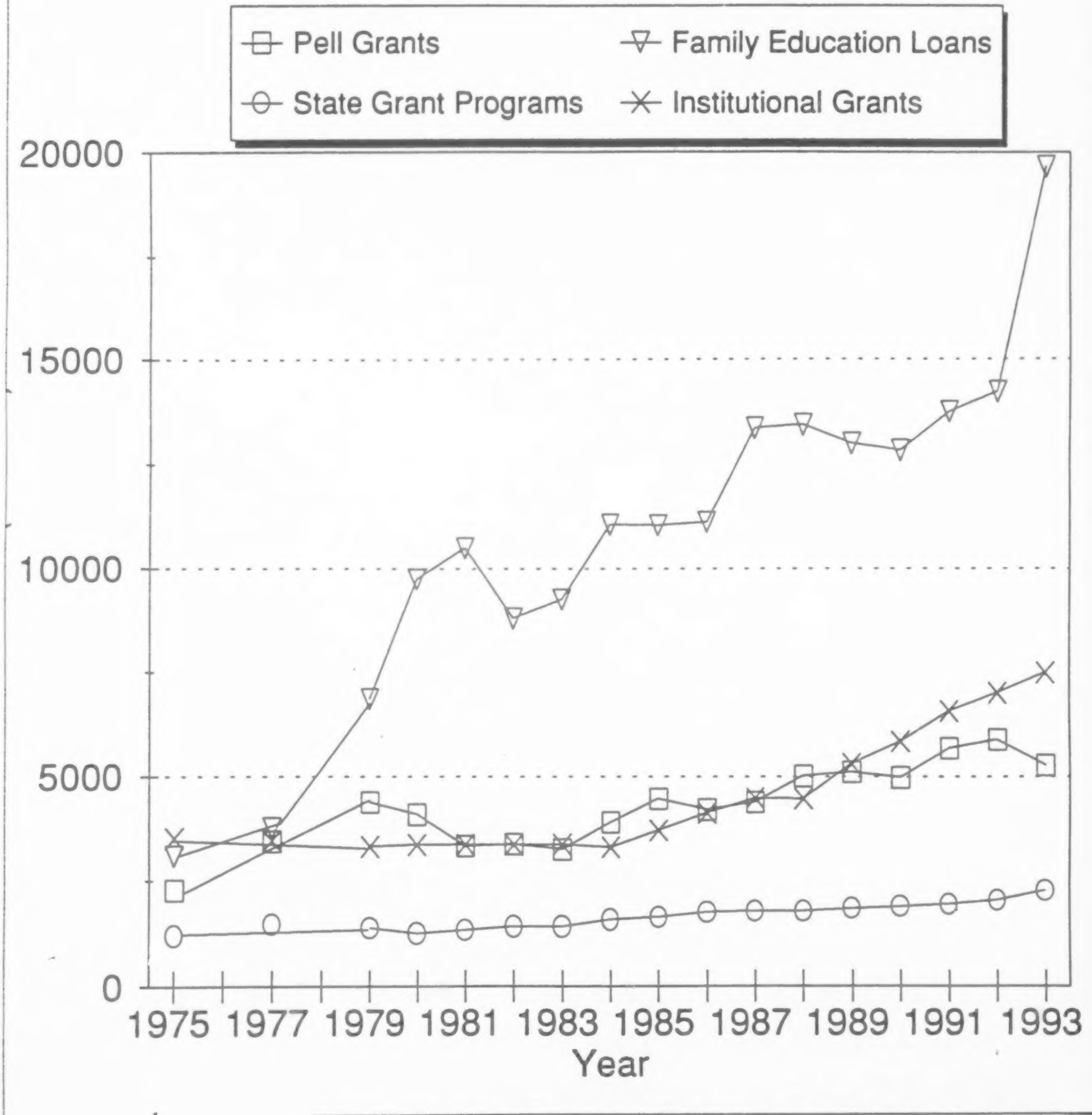
Enrollment at Age 18-19 and Reported College Entry at Age 21 by Cohort

Figure 5: Student Aid and Tuition/Fees



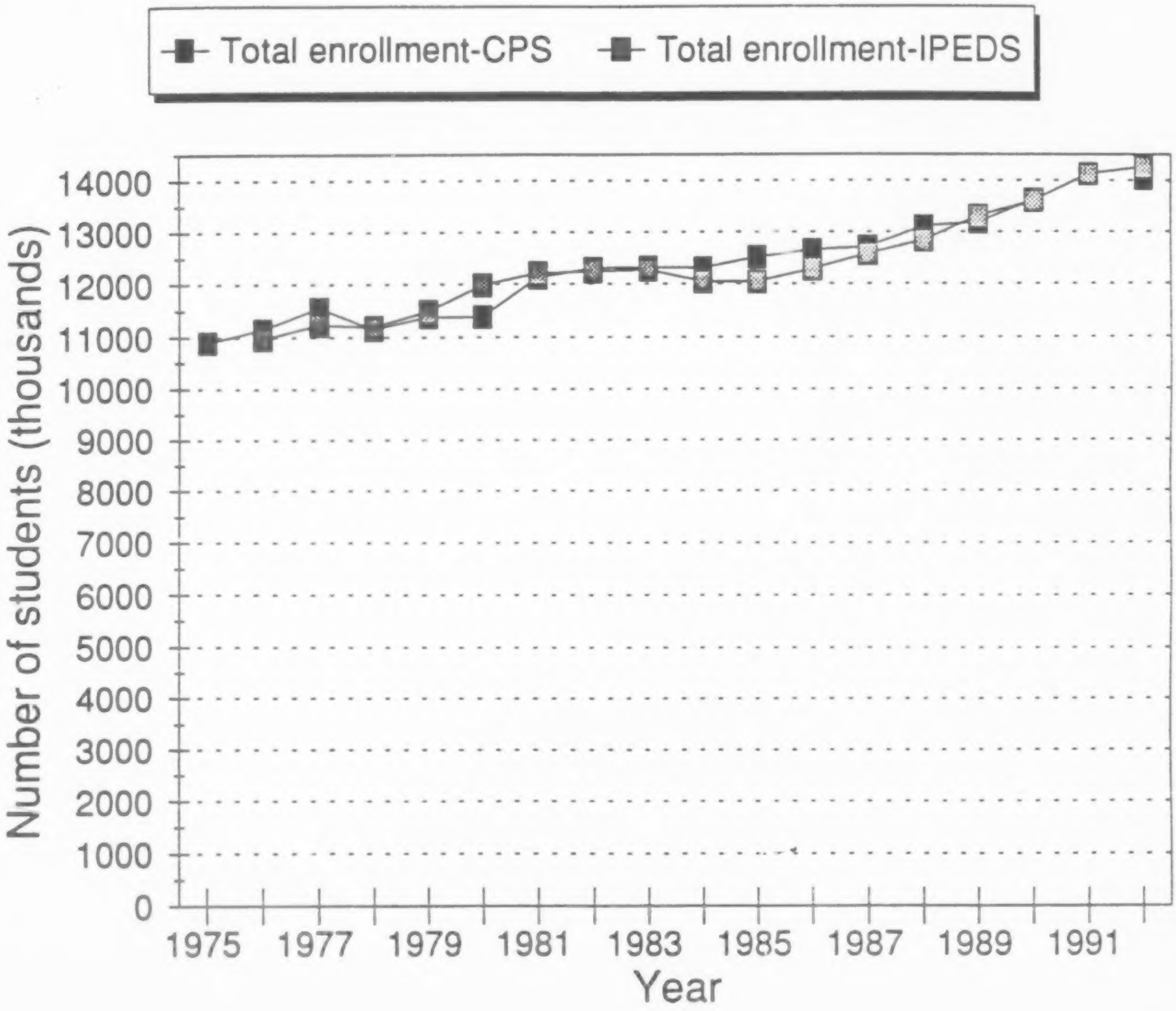
Source: "Digest of Educaiton Statistics 1994," Tables 304, 331, 332 and 333. All numbers in 1991 CPI dollars. All institutional aid except Pell Grants included.

Figure 6: Student Aid by Source

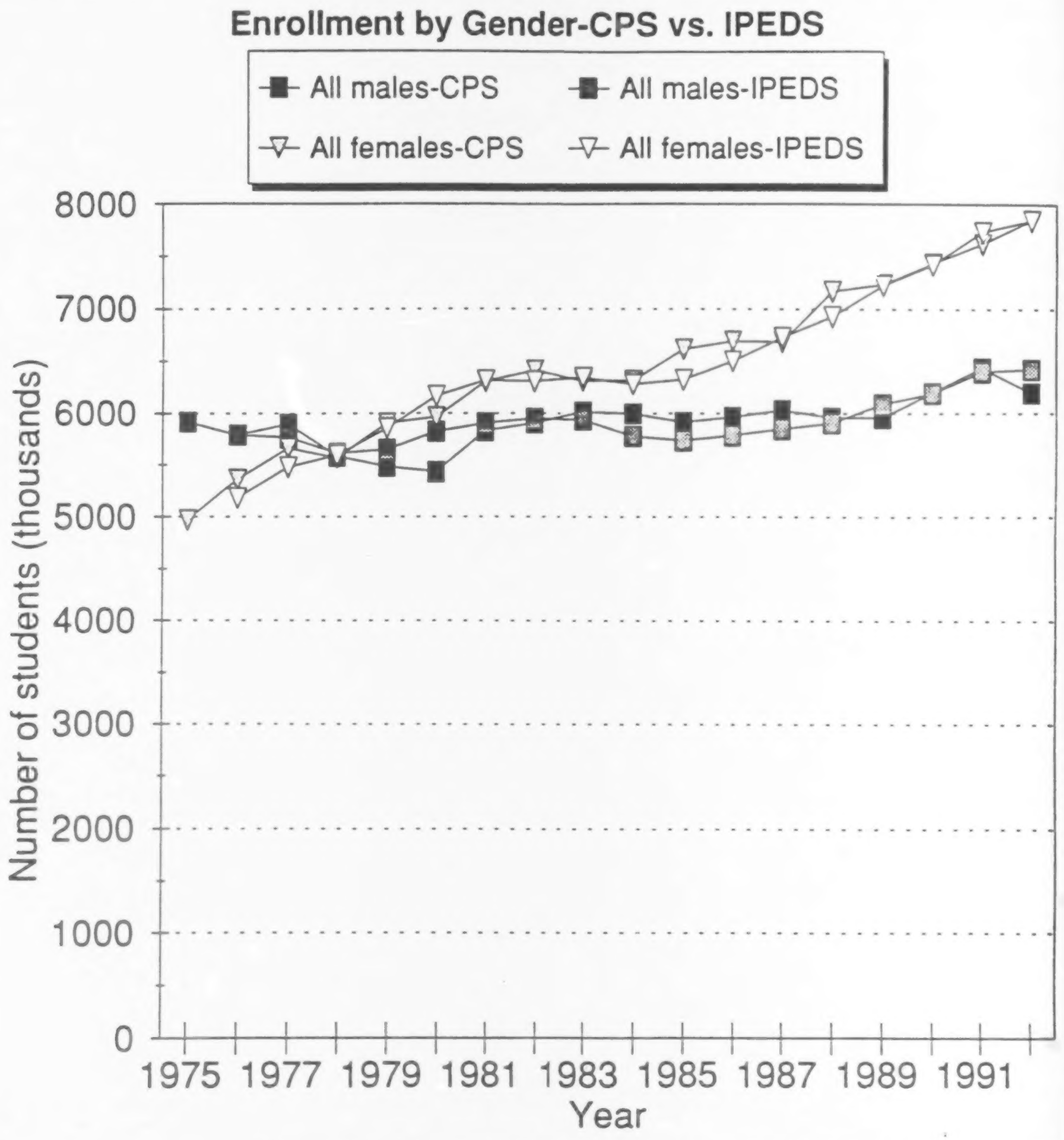


Source: "Trends in Student Aid: 1963 to 1983," and "Trends in Student Aid: 1984 to 1994," Table A-2 and Table 2, respectively, The College Board. Figures for 1981-1983 and 1992-1993 are based on estimates. Family Educational Loans include GSL, PLUS and SLS.

Total Enrollment - CPS vs. IPEDS



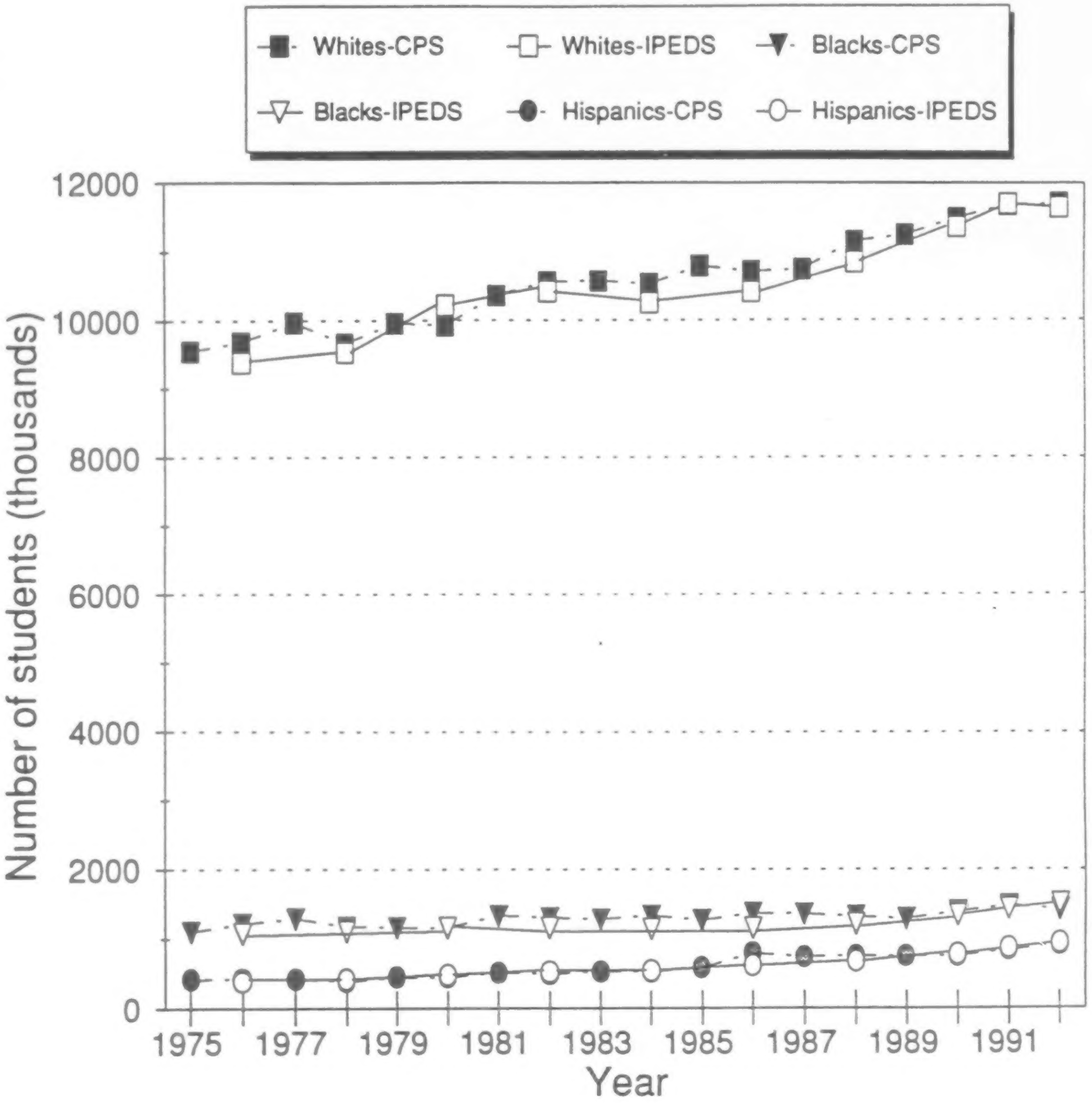
Source: "School Enrollment-Social and Economic Characteristics of Students: Oct. 1992," Table A-5, Series P-20, Bureau of the Census and "Digest of Education Statistics," Table 169, National Center for Education Statistics (total students-proprietary students)..



Enrollment by Gender-CPS vs. IPEDS

Source: "School Enrollment and Economic Charateristics of Students:Oct 1992," Table A5, Series P-20, Bureau of the Census and "Digest of Education Statistics," Table 169 from 1994, National Center for Education Statistics. Calculation from Table 169: total male or female students minus estimated number of male or females in proprietary schools.

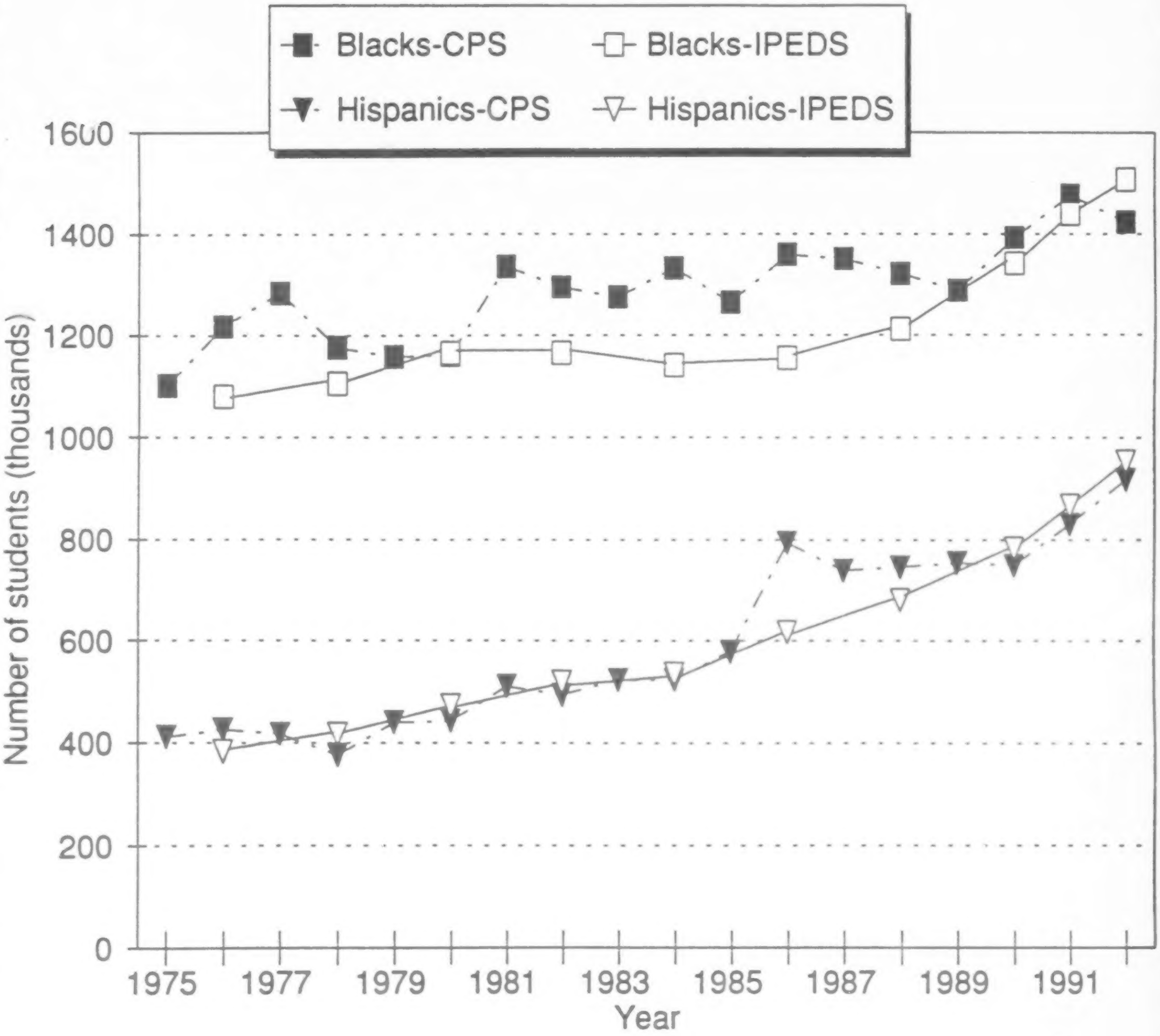
Enrollment by Race-CPS vs. IPEDS



Source: "School Enrollment-Social and Economic Characteristics of Students:Oct 1992," Table A5, Series P-20, Bureau of the Census and "Digest of Educational Statistics," Table 203 from 1994 and Table 174 from 1990, National Center for Education Statistics.Note: No IPEDS data on race for odd years until 1991. 81% of Hispanics assumed to be white, 12% assumed to be black.

Enrollment of Blacks and Hispanics

CPS vs. IPEDS



Source: "School Enrollment-Social and Economic Characteristics of Students:Oct 1992," Table A5, Series P-20, Bureau of the Census and "Digest of Educational Statistics," Table 203 from 1994 and Table 174 from 1990, National Center for Education Statistics. Note: 81% hispanics assumed to be white, 12% assumed to be black. Also, IPED sata available every other year until 1990.

Indicators of Educational Achievement

Daniel Koretz
Resident Scholar
RAND Institute on Education and Training

January, 1995

Paper prepared for a conference on "Indicators of Children's Well-Being," Washington, D.C., November 17-18, 1994, sponsored by the Institute for Research on Poverty at the University of Wisconsin-Madison, The Office of the Assistant Secretary for Planning and Evaluation of the Department of Health and Human Services, the National Institute on Child Health and Human Development, and Child Trends, Inc. The views in this paper are solely those of the author and do not represent the position of RAND or of the conference sponsors.

Indicators of Educational Achievement

Conventionally, "educational achievement" is used in social science as to mean mastery of knowledge and skills or, more narrowly, performance on specific tests of knowledge and skills. Thus narrowly defined, achievement stands in contrast to "attainment," which typically is used to refer to the levels of schooling individuals complete. In keeping with this traditional if somewhat arbitrary usage, this paper uses "indicators of educational achievement" to refer to some classes of educational tests--or, as it is now more fashionable to say, "assessments." The paper considers recent trends in the uses made of achievement tests; characteristics of available achievement measures; limitations of the measures; issues that arise in building a system of achievement indicators; and some possible steps toward a stronger indicator system.

To understand achievement indicators, it is helpful to contrast them to other indicators of children's status, or to other social indicators more generally. Over the past decade and a half, I have worked intensively with diverse social indicators, including postsecondary enrollment rates, dropout rates, measures of progress through school, poverty rates, health-care utilization rates, and incidence and prevalence rates for various mental illnesses. All pose vexing and, on better days, fascinating challenges pertaining to data collection, incomplete or missing data, operationalization of constructs, choice of metrics, choice of analytical frameworks, and so on. In many respects, however, achievement indicators stand out as particularly difficult. For a variety of reasons, achievement data tend to be sparser, less robust, and more expensive to obtain than data in many other areas, and they are routinely--indeed, systematically--misinterpreted by many of the key audiences for which they are produced. Moreover, at the present time, achievement data are frequently required to serve several distinct and even conflicting functions. As a consequence, the value of these data as indicators is degraded, and improvement of the indicator system is made more difficult. Some of the bases for these conclusions are discussed later in this paper.

In the following section, I will comment on the characteristics of indicators. I will then discuss current achievement measures and their limitations. In conclusion, I will discuss some of the implications for the construction of a better system of achievement indicators.

SOME CHARACTERISTICS OF INDICATORS

The characteristics of educational indicators have been the subject of extensive writings (e.g., Raizen and Jones, 1985; Murnane and Raizen, 1988; Shavelson, McDonnell, Oakes, and Carey, 1987), and to summarize that literature would go well beyond the scope of this paper. However, for present purposes, it is important to note some important attributes shared by most educational indicators.

Perhaps most important, the basic function of indicators is *descriptive*--to describe, for example, the income distribution of households with children, the health status of impoverished preschoolers, the range of science course offerings in high schools, or the mean mathematics achievement of American 8th-graders. That does not mean that appropriate uses of indicators are restricted to simple univariate and bivariate statistics. However, it does mean that databases that are well designed, or at least efficiently designed, to provide indicators will often be poorly designed or even totally inadequate to support causal inferences.

Indicators are intended to provide descriptive information at various levels of aggregation, such as poverty rates among school-age children or relative trends in achievement among racial/ethnic groups. Data that are best suited to providing aggregate information are often not the best for providing individual data, and indeed, different levels of aggregation often call for different design decisions. As discussed below, these design conflicts are acute in the case of achievement data.

Indicators are typically used to support very general, broad conclusions (Koretz, 1992b). For example, for many purposes, one needs to know variations in price changes across categories of goods, but the CPI is not able to tell one that; it simply provides an overall index of inflation. Similarly, the poverty rate tells one nothing about changes in the income sources of poor people, but it does provide a useful if crude picture of trends in the overall incidence of poverty. Achievement indicators are no exception to this generalization. In using indicators, the public is interested in constructs such as "mastery of secondary school mathematics" or "mastery of the mathematics needed for technologically advanced occupations," not "mastery of quadratic equations" or even, in most cases, "mastery of elementary algebra." To be a suitable basis for indicators, an achievement test should be built with these broad inferences in mind.

Because of the broad conclusions that indicators are used to support, there is a pervasive tension in building indicators between comprehensiveness and simplicity of reported data. Moreover, different approaches to simplification of the data will often produce apparently different answers. For example, trends in the *numbers* of minority and white youth enrolled in postsecondary education show somewhat different trends than do enrollment *rates*, because trends in the size of the cohorts--the denominator in the rates--vary across racial/ethnic groups (Koretz, 1990). Such considerations, which are arguably more severe in the case of achievement indicators than many other types of indicators, are generally obscure to the lay audiences that are among the most important consumers of indicators.

Finally, unless they can be obtained from administrative data, indicators are most often drawn from broadly based, multi-purpose social surveys. This is primarily because of the need for large, representative samples. One consequence, however, is that information pertaining to smaller groups or certain specific topics may be inadequate or totally lacking. This is particularly important in the case of achievement indicators, because they require sampling of tasks as well as sampling of individuals.

FUNCTIONS OF ACHIEVEMENT MEASURES

Achievement tests are currently used to serve three functions: *individual measurement; monitoring of groups, schools, or systems;* and *accountability*. Indicator data, including achievement indicators, are aggregate statistics used to describe the output of the educational system or its components and thus are one instance of the second of these functions. The functions are not entirely distinct in the real world. In particular, one reason policymakers want to monitor schools is to hold educators accountable. Nonetheless, these functions are to some degree inconsistent; in particular, some uses undermine the utility of test data as indicators.

Standardized achievement testing--meaning achievement testing in which tasks, administrative conditions, and scoring are made uniform--has a long history in the United States. Resnick (1982) identified the use of tests with published directions and uniform scoring and interpretation as early as the 1840s. The first standardized achievement test battery, the Stanford Achievement Test, was initially published in 1923 (Resnick, 1982). The role of standardized testing grew markedly if erratically over the following years (for example, during the 1930s [Haney, 1981]) and has been a staple of elementary and secondary education for decades.

The amount and purposes of achievement testing have changed dramatically over the past several decades (see Koretz, 1992a). Although monitoring and accountability both provided an early impetus for achievement testing (Haney, 1984), neither of these functions were salient for the first two decades after World War II. Rather, during those years, standardized tests were used primarily to assess individual students, and to a lesser degree to evaluate curricula (Goslin, 1963, and Goslin, Epstein, and Hallock, 1965). Generally, tests were "low-stakes"--that is, the consequences of test scores were minor for most teachers and students.

The functions of achievement testing began to change in the second half of the 1960s. One critically important change was the enactment of the Elementary and Secondary Achievement Act (ESEA). Title I of ESEA authorized the federal compensatory education program and established achievement testing as a primary mechanism for monitoring and evaluating it. Title I programs were eventually established in the overwhelming majority of school districts, and its testing programs are widely considered to have had a seminal influence on testing throughout the elementary and secondary education system (e.g., Airasian, 1987; Roeber, 1988). A second milestone was the establishment of the National Assessment of Educational Progress (NAEP), a recurring assessment of the achievement of a nationally representative sample of youth. The NAEP was originally intended solely as an indicator of achievement--i.e., a source of descriptive information. Because is focus was all youth, it originally sampled not students in given grades, but rather individuals of given ages, in or out of school. By virtue of its frequency, representative sampling, and broad content coverage, NAEP rapidly became the preeminent national indicator educational achievement. As discussed below, its functions are now changing, with substantial implications for a national indicator system.

The next step in the transformation of achievement testing was the growth of state-mandated "minimum-competency testing" in the 1970s (see Jaeger, 1982). These testing programs aimed not merely to measure performance, but also to improve it. They did so by imposing serious consequences on students for failure. The specific consequences varied; a majority of those implemented by the early 1980s were used as "exit exams," (to set minimum standards for graduation from high school), while a smaller number were used as "promotional gates" (to determine eligibility for promotion between grades.

Close on the heels of the minimum-competency testing movement came the "educational reform movement" of the 1980s. Early in that decade, policymakers and the public became increasingly aware of weaknesses in the performance of American students. Debate about the decline in aggregate test scores that had begun in the mid-1960s (and which in fact had already ended; see Koretz, 1986) belatedly intensified. NAEP revealed that many students were failing to master even rudimentary skills, and a number of studies showed that American students compared unfavorably to their peers in other nations. Growing concern was expressed in a spate of commission reports, the most important of which was probably *A Nation at Risk* (National Commission on Excellence in Education, 1983).

This concern spawned a wave of reform policies, particularly at the state level, the most consistent theme of which was the growing use of standardized tests as accountability devices. For example, Pipho (1985) noted that "Nearly every large education reform effort of the past few years has either mandated a new form of testing or expanded uses of existing testing." This new testing was generally tied to serious consequences for students, educators, or entire school systems (Koretz, 1992a). Overall, however, the reform movement showed a shift away from stakes for students and towards evaluations of entire schools or systems.

Evaluations of the testing programs of the reform movement are few, but the available evidence suggested that they could produce inflated test scores and degraded instruction (e.g., Koretz, Linn, Dunbar, and Shepard, 1991; Shepard and Dougherty, 1991). The programs rapidly fell into disfavor in the policy world and were replaced by a "second wave" of education reform that continues at present. The diverse programs of the second wave typically continue the reliance on accountability-oriented testing as the prime engine of educational improvement and the shift toward aggregate accountability and consequences for educators rather than only students. The current programs differ from their predecessors, however, in placing less reliance (in some cases, none at all) on traditional multiple-choice tests, using instead diverse "performance assessments." (The phrase is ubiquitous but has no common definition; it includes virtually anything that is not multiple-choice, including short constructed response questions, essays, portfolios, large-scale performance events, and group projects.) An archetype is the Kentucky Education Reform Act, or KERA. KERA established a school performance index that comprises both achievement tests and non-cognitive measures, such as dropout rates. The tests, which are given by far the greatest weight in the index, originally included both multiple-choice tests and performance assessments, but the former were dropped recently. KERA imposes a range of substantial financial rewards and serious sanctions that will be assigned to schools on the basis of the amount of change they show on the index. Other states, such as Vermont, have instituted lower-stakes performance assessment systems in which the publication of district or school scores is expected to be sufficient source of pressure.

For present purposes, an essential aspect of the testing programs of the 1980s and 1990s is that they use test scores for a wide range of purposes. Perhaps most important, they typically use the same tests to hold individuals accountable and as indicators of progress. These roles of accountability and monitoring conflict because the latter can corrupt scores and those undermine the value of scores as indicators. (This is discussed further below in the section on corruption from behavioral response.)

Another pertinent conflict is between aggregate monitoring and student-level feedback. Because of the broad inferences that are typically based on achievement indicators--e.g., conclusions about changes in the mean "mathematics proficiency" of the nation's high school seniors--most experts agree that an assessment used as an aggregate-level indicator should be matrix-sampled. A matrix-sampled assessment includes far more items than an individual student can take, and each student is given a systematic sample of the total item pool. This approach, used in NAEP, increases the breadth and utility of the assessment as an indicator (and also decreases the probability of teaching to the test that might corrupt the scores). At the same time, matrix sampling makes it far harder (sometimes impossible) to obtain reliable or comparable scores for individual students. In state after state, however, policymakers, educators, and the public have expressed dissatisfaction with the resulting lack of student-level scores. (This was one reason why the governor of California recently terminated the state's nationally renowned performance assessment program.) Assessments can be redesigned to provide student-level scores, but often at the cost of lessening the quality of the data for aggregate monitoring.

The following two sections, which discuss the characteristics of achievement measures and some of their limitations, should clarify some of the reasons why the uses of assessments often conflict.

THE NATURE OF TESTS

Lay observers often treat achievement-test scores as synonymous with achievement. But in almost all cases of interest, achievement is a latent variable, and tests are an incomplete measure of it. Tests are sets of tasks designed to elicit behaviors reflecting that latent variable. Those sets of tasks are generally small samples of the "domain" of achievement they are designed to represent. Moreover, even if they are representative in terms of some ideal mix of content and skills, they are often somewhat unrepresentative in terms of formats and context, because some formats and contexts are difficult to include in tests. These sampling characteristics hold the key to many of the limitations of achievement indicators.

To illustrate the magnitude of this sampling problem, consider three domains of achievement: writing mechanics, vocabulary, and mathematics at grade 12. The skills comprising writing mechanics are quite few. There are a limited number of rules governing, for example, punctuation or capitalization. Therefore, a test of writing mechanics can be a reasonably large sample of the relevant domain of knowledge and skills. At the other extreme, consider vocabulary. Studies indicate that reasonable fluency in most natural languages requires knowledge of literally thousands of words. Even adolescents, whose productive vocabulary often seems to be two orders of magnitude smaller, comprehend thousands of words. No one, however, is prepared to construct a test comprising thousands of words, and no one would spend the time and money required to administer one if it were constructed. Hence tests of vocabulary are typically small--say, 40 or 60 words. A test in a subject area such as mathematics may be less extreme than vocabulary, but it still represents a small sample. For example, the grade 12 NAEP test is cumulative; that is, it is intended to assess the range of mathematics skills that students are expected to have acquire by that point in their schooling, from simple arithmetic on up. Moreover, the NAEP is a broader test (in terms of the number and range of its items) than many others. Yet the NAEP grade 12 mathematics test in 1992 comprised only 179 items--an average of 15 items per year of schooling.[1]

The consequence of this limited sampling is that *scores on a test are only meaningful to the extent that one can generalize from them to mastery of the broader domain the test is used to represent.* No one cares whether some students know 30 of the specific words on a vocabulary test, while others know 35. Rather, people care about differences in working vocabulary that are revealed by those differences in test performance. Similarly, when using NAEP as an indicator, few are interested in the specific items included in the tests, except to the extent that the items are illustrative. Rather, most observers are interested in what they can infer about students' mastery of elementary and secondary mathematics.

---

[1] These counts include only the items in the main NAEP assessment that were scaled.

The current movement toward performance assessment has several ramifications for the adequacy of task sampling. On the positive side, this trend may result in reliance on a broader and less unrepresentative sampling of formats. To take a particularly uncontroversial example, adding direct tests of students' ability to write provides a more representative sampling of skills pertaining to writing one would have from a multiple-choice test of language mechanics taken alone. On the negative side, greater reliance on performance assessment will likely severely exacerbate the problem of sampling of skills and knowledge, for two reasons. First, performance assessment tasks typically take much more time, so students can complete far fewer tasks per unit of time. Second, by virtue of their complexity, scores on performance tasks typically include a substantial amount of "task-specific variance." That is, scores are substantially affected by idiosyncratic characteristics of the tasks. The consequence is that performance generally correlates poorly across theoretically related performance tasks (e.g., Dunbar, Koretz, and Hoover, 1991; Shavelson, Baxter, and Gao, 1993). This is true even of essay tests of writing, which would appear on their face to have less task-specific variance than many current assessments in areas such as science, some of which entail hands-on work with materials and apparatus. Moreover, there is evidence that much of this task-specific variance is "construct-irrelevant"--that is, largely or totally irrelevant to the latent variable of interest. Thus, performance assessment generally will increase the amount of time (and money) required to obtain a reasonable sample of a domain of interest. Wainer and Thissen (1993), in an analysis of open-ended and multiple-choice questions on Advanced Placement exams, humorously suggested quantifying this using measures he labeled "reliamin" and "reliabuck"--unit reliability per minute or dollar--and concluded that even essay questions appreciably lower an examination's value on both measures. Therefore, performance assessment increases the pressure to rely on matrix-sampled assessments and increases the tension between providing aggregate data suitable for indicator use and individual-level data suitable for diagnosis or student-level accountability.

LIMITATIONS OF ACHIEVEMENT TESTS

Because of their characteristics, tests have a number of important limitations as measures of the latent construct of children's achievement. In this section, I will briefly describe three limitations that should be of central concern in building an indicator system.

The Problem of Limited Robustness Across Measures

Because tests are generally sparse samples from large and varied domains, results often vary among theoretically comparable measures. To put this variation into perspective, the simple correlations among well-built tests of a given domain are generally very high.[2] Despite these high correlations, however, results often do differ in important and often unanticipated ways from one test to another, particularly when statistics other than rankings of individual students are at issue.

One particularly tidy example of the lack of robustness across measures comes from the historical data maintained by the Iowa Testing Programs at the University of Iowa, which are the longest time series of internally consistent achievement data available for students in the United States. Almost all students in Iowa take the Iowa Tests of Basic Skills through grade 8 and the Iowa Test of Educational Development in grades 9 through 11. Thus, the ITED trends for grade 9 represent almost exactly the same cohorts of students as ITBS trends for grade 8, but lagging by one year. Yet during the period of declining test scores (roughly the mid-1960s to the late 1970s), the mean scores of Iowa eighth-graders on the ITBS dropped roughly twice as much as did the scores of the same students on the ITED in grade 9 (Koretz, 1986, pp. 53-54).

Another, more timely example can be found in a study conducted by Linn, Kiplinger, Chapman, and LeMahieu (1992) for the New Standards Project, a huge national effort to develop new performance assessments that will shape instruction. Linn *et al* examined the impact of having writing assessments scored by raters from different states. That is, students in each state wrote essays in keeping with the requirements of their own state assessments, and raters from that state used the assessment's scoring rubrics. Papers were also given to raters from other states, however, and these raters applied the scoring rubrics from *their* states rather than those from the students' own states. This was an attempt to discern whether locally developed rubrics can be made to reflect a common set of underlying standards. Linn, *et al.*. found that correlations were very high across sets of raters, but means were sometimes substantially discrepant. That is, raters from all states ranked students similarly, but they often differed in assigning levels of performance to them.

---

[2] In fact, the correlations among tests of *different* domains are typically sizable as well. For present purposes, however, it is not necessary to go into the long-standing arguments about the meaning of these cross-subject correlations or of the general factor they imply.

The Problem of Limited Robustness across Metrics

Another limitation of robustness is variation in results across alternative metrics. This variation can stem from the characteristics of the test itself, the choice of scale (and scaling method), or characteristics of the distribution of scores.

More than most indicators, achievement tests lack a "natural" metric. By way of contrast, consider dropout rates. There are numerous ways to tabulate dropout rates, including the percentage of 10th graders who fail to graduate on schedule (used in HSB and NELS) and the proportion of an age group (typically 16 to 24 or 18 to 24) who are not enrolled and have not graduated (see National Center for Education Statistics, 1988). These two metrics mean different things and provide very different pictures of dropout rates, both cross-sectionally and over time. But both are variants of a single metric: a rate, with dropouts (variously defined) in the numerator and a relevant base cohort (variously defined) in the denominator.

The closest analog for an achievement test would be the percentage of items answered correctly--the standard measure for the scores of teacher-created tests that we all took during school. This ratio, however, has no clear meaning, because it is entirely a function of the specific items included on the test. For that reason, most large-scale assessments, such as commercial achievement tests and the National Assessment (for the past decade or so) avoid the percentage-correct measure and use instead one or more types of scaled scores. The commonly used scales, however, are not linear transformations of teach other, and the conclusions one reaches can depend on the scale employed. Choice of scale can influence, for example, conclusions about relative trends in high-and low-achieving groups (e.g., Spencer, 1983; Koretz, 1986) and about changes in the variability of scores as students progress through school (Clemans, 1993; Burket, 1984; Hoover, 1984a, 1984b, 1988; Yen, 1988).

Even once a scale has been chosen, achievement indicators can present different pictures depending on which aspect of the distribution of scores is the focus of attention. For example, because the variance of test scores differs among states, the National assessment shows that states rank differently in terms of means than in terms of their 75th percentiles (Linn, Shepard, and Hartka, 1992; Mullis, Dossey, Owen, and Phillips, 1991, 1993).

A lack of robustness due to metric shows up clearly in a class of indicators that could be called "representational indicators"--that is, indicators of the representation of different groups in specified ranges of the achievement distribution. An example would be the proportion of each racial/ethnic group falling into the top quartile or top decile of the distribution. In general, if scores are approximately normal and have similar variances among groups, representational indicators will show progressively more extreme under-representation of low-achieving groups as the threshold for the indicator is raised. For example, the under-representation of African American students will typically be more severe in the top decile than in the top quartile. We have found that pattern in NAEP (Koretz and Lewis, n.d.), and Hedges (personal communication) has found it in the NLS-Y. This pattern has occasionally been noticed in the lay world and has sometimes been cited as an indication that the educational system fails high-achieving minority students more than it fails minority students in general. While this conclusion may or may not be correct, it does not follow from the simple fact of greater under-representation at high levels of achievement. Rather, given certain common distributional characteristics, that pattern can stem from nothing more than a mean difference between groups.

The Problem of Corruption from Behavioral Response

Perhaps the single most vexing aspect of achievement tests for purposes of building indicators is their susceptibility to corruption from behavioral responses to measurement. This problem is not limited to achievement tests, of course. The nominal budget deficit under Gramm-Rudman-Hollings is an example of a corruptible measure. Those who follow budgetary matters have frequently seen the Congress doing things like moving military pay days backwards or forwards to cross the magical October 1 deadline. This changes the official deficit figure for the target fiscal year but of course has no real effect on the latent variable of interest, the actual budget deficit. People who travel frequently are familiar with another example: on-time statistics. Airline flights are frequently on time, now that public attention is focused on that particular indicator, but this level of service has been obtained in part by setting arrival times far later than actual flight time requires.

Achievement measures, however, are particularly susceptible to corruption. The mechanism of this corruption is inappropriate teaching to the test, also commonly called "coaching" or (outside the United States) "cramming." The dividing line between inappropriate and appropriate teaching to the test is hazy and the subject of intense debate, and address it adequately would go beyond the scope of this paper. The basic issue, however, is that if teaching is narrowed to focus on the specific content of the test, the test will become less representative as a sample of the achievement domain of interest. Scores than become inflated as measures of mastery of that domain. Quantitative research on the inflation of test scores is limited, but there is evidence that it can be large. One study found inflation of mathematics scores of roughly half an academic year by the spring of grade 3 (Koretz, Linn, Dunbar, and Shepard, 1991).

The problem of inflated test scores indicates a fundamental conflict between two of the functions often assigned to large-scale assessments: accountability and monitoring. Accountability will typically induce teaching to the test. Steps can be taken (for example, use of a very broad, matrix sampled test) to lessen inappropriately narrowed instruction and its deleterious effects on scores. However, it is unrealistic to expect that an accountability-oriented testing program can entirely avoid the inflation of scores. In contrast, accurate monitoring of aggregate trends in achievement--that is, maintenance of high-quality indicators--requires that achievement measures remain unsullied.

ISSUES IN BUILDING ACHIEVEMENT INDICATORS

A principal focus of this conference is steps that should be taken to improve indicators of children's well-being. This section notes several recommendations for achievement indicators, some of which stem directly from the characteristics of achievement tests noted above.

Maintaining Multiple Indicators

Although "multiple measures" has become a mantra in policy circles, it is a principle most often observed in the breach. Multiple measures are important throughout an indicator system, because reliance on a single measure or a single database leaves the risk that substantive findings will be confounded with measurement effects stemming from the peculiarities of a sample, the specifics of a survey instrument, the operationalization of constructs, etc. Use of multiple measures is particularly important for achievement measures, however, because of the sampling of tasks and the concomitant risk of variation among measures. As I will note below, the national indicator system currently includes too sparse a set of achievement measures.

Deciding on Levels of Aggregation

Existing national, state, and local assessment programs vary in the levels of aggregation at which they provide data. Various programs provide data at the levels of students, policy-relevant groups of students (such as racial and ethnic minorities), classrooms (or teachers), schools, local education agencies, intermediate education agencies, states, and regions. Currently, there is a clear trend at both the state and national levels to focus accountability pressures at the levels of schools and teachers. This is evident, for example, in the Title I provisions in the recent reauthorization of ESEA and in state education reforms in Kentucky, Maryland, Vermont, and elsewhere. This emphasis is reflected in the design of state assessment programs, which are increasingly employing matrix-sampled designs to obtain school- and district-level statistics.

The optimal levels of aggregation of course depend on the purposes of the assessment program. The goal of institutional accountability clearly suggests the need to maximize the quality of estimates at the levels of aggregation at which educational decisions are made. However, it is important to note that aggregate statistics may be important even if assessment data are to be used only for the descriptive purposes that are most appropriate for indicators. The National Assessment provides an example. Originally, NAEP was designed to provide only estimates at very high levels of aggregation: the nation as a whole, regions, racial/ethnic groups, and so on. The NAEP sample design reflected this; samples within states were typically insufficient to provide reliable state-level estimates, and the samples in many schools were far too small to provide reliable school-level estimates. In the late 1980s, however, pressure for reliable state-level estimates increased, and Congress authorized additional NAEP samples to provide them. Currently, there is widespread interest in school-level descriptive information, such as differences in the performance of students in schools with high or low mean socioeconomic status (SES) or between students in certain types of classes. However, the sample design has not yet been altered in ways (such as imposition of minimum within-school sample sizes or adding classes as a stratum for sampling) that would support such statistics well.

Unfortunately, there are trade-offs among levels of aggregation that are often unrecognized in the policy world, and the priorities among possible uses of the data is often a matter of disagreement. The trade-offs among levels of aggregation are of two types in the case of achievement indicators. Trade-offs involving *sampling of individuals* are similar to those pertaining to other indicators. For example, if NAEP were to increase within-school samples or introduce classroom-level sampling, the result would be a more highly clustered sample that is less efficient for estimates at higher levels of aggregation. This lower efficiency could be offset only by increasing the total sample and thus the total cost of the survey. But achievement indicators are also affected by measurement trade-offs involving *sampling of tasks* that do not come into play with many other types of indicators. For example, an assessment designed to provide high-quality aggregate estimates will often be matrix-sampled; this increases the breadth of the assessment and its efficiency for producing aggregate estimates but at the cost of limiting or even precluding scores for individual students. Moreover, different levels of aggregation suggest different matrix-sampling designs. Thus the quality of the assessment for providing information at any given level of aggregation depends largely on design decisions about sampling of both individuals and tasks that are made before the assessment is fielded.

Reporting by Subgroups

National achievement data are generally reported for a variety of subgroups. The National Assessment, for example, has traditionally reported scores by race/ethnicity, region, parental education, and a much criticized composite urbanicity/SES measure. (This composite, called "size and type of community," classifies schools as "advantaged urban," "disadvantaged urban," "extreme rural," etc., based on type of community and principals' estimates of the occupational profiles of parents.) State and local data are often reported for fewer subgroups; for example, some jurisdictions don't report separately for racial/ethnic subgroups.

Several limitations of the reporting of achievement by subgroups should be noted. First, reporting tends to be based on *a priori* and conventional classifications, sometimes without clear agreement about purposes. In the case of some subgroups that are of clear importance to policy (such as racial/ethnic breakdowns), this has not been problematic, but for other groupings, it has been. For example, there has substantial debate about the background variables that should be collected by and use for the reporting of NAEP. That debate hinges on decisions not yet made about the most important purposes of the reporting by subgroups. For example, some observers would like NAEP to focus its measures of SES, but if NAEP is to report trends for schools facing particularly severe educational challenges, variables not typically considered part of SES, such as proficiency in English, may be essential.

Second, most achievement databases lack certain background variables that are potentially important for reporting. Few achievement databases, for example, include trustworthy data about household income. State and local data generally have no household income data at all, as local and state agencies are not entitled to ask for that information. (Variables such as the percentage of students receiving free or reduced-price lunch are often used as poor proxies for poverty rates.) Even the National Assessment has no income data, because it does not include a parent survey. Only occasional research-oriented data collection efforts provide reasonable income data. The few representative databases that provide good measures poverty over time, such as the Survey of Program Participation or the Panel Study on Income Dynamics, lack achievement measures. Most achievement databases also lack information about the place of students' and parents' birth. At a time of rapid immigration, this is a serious shortcoming; it precludes not only reporting for immigrants, but also disentangling the achievement of new immigrants from trends in the performance of native-born students in the same racial/ethnic group. (For example, some observers suspect that the lack of consistent progress in the status drop-out rate of Hispanics--in contrast to blacks--may reflect an influx of dropout-prone poor immigrants that obscures progress shown by native-born Hispanics.)

Third, achievement databases lack sufficient samples to report for certain potentially important subgroups. For example, for some purposes, it may be sufficient to treat "Hispanic" as a single category, but the limited available evidence suggests that patterns of achievement and attainment differ substantially among Hispanic ethnic groups. Similarly, it would be naive to expect poor immigrants from Southeast Asia to perform similarly to the ethnically different native-born Asian-Americans, who typically show mean scores higher than those of Anglo students in some subject areas. Specific Hispanic or Asian subgroups, however, are mostly quite small, and a sampling rate that would provide reliable estimates for them will generally be prohibitive.

Adjusting Indicators for Compositional Differences

Achievement indicators are substantially influenced by adjustment for demographic differences. Adjustment has large effects on cross-sectional comparisons because of the typically large mean differences in scores between certain groups, particularly racial/ethnic groups. Time series are affected by adjustment not only because of differential trends among groups (such as the now well-recognized relative gains of black students), but also because of immigration and group differences in fertility.

Whether achievement indicators should be adjusted for compositional differences, however, is currently a matter of intense debate in the policy community. On the one hand, it is widely recognized that failure to adjust for compositional differences will undermine the fairness of cross-sectional comparisons. On the other hand, many in the policy world argue that adjusting scores for factors such as differences in racial composition or poverty rates reifies current disparities in performance and undermines the currently widespread push for higher standards for all students. These arguments, for example, were recently raised in a dispute about whether National Assessment results should be adjusted to make comparisons among states "fair." Not surprisingly, research shows that state rankings are substantially influenced by differences in the composition of the student body (e.g., Linn Shepard, and Hartka, 1992). However, the decision of the National Assessment Governing Board was that scores should *not* be adjusted.

The common arguments against adjustment lose much of their force when the issue is trends in achievement rather than cross-sectional comparisons. Yet, oddly enough, changes in the composition of the student body are rarely taken into account when trends in achievement are discussed. Bracey is an exception; he has argued that trends in achievement represent a striking success when changes in the demographic composition of the student population are considered (e.g., Bracey, 1991). Bracey did not actually estimate the effects of demographic changes, however, and his assertion is overstated: demographic change can account for only a modest portion of the pervasive decline in achievement that occurred during the 1960s and 1970s (Koretz, 1987, 1992c).

The importance of taking compositional changes into account is illustrated by recent trends in the SAT. The familiar trends in overall mean SAT scores are shown in Figure 1, expressed as differences in standard deviations from the low point of 1980. Several years ago, Jaeger (1992) pointed out that the lack of improvement in the grand mean was attributable to compositional changes rather than a lack of progress within racial/ethnic groups. This can be seen, albeit not too clearly because of the large number of groups, in Figure

2. The grand mean in mathematics was only 2 points higher in 1993 than in 1976--a difference that is small both substantively and in comparison to the fluctuations in the mean during the intervening years. During that period, however, the mean mathematics score in every racial/ethnic group other than the "other Hispanic" group went more than the grand mean. The means for Mexican-Americans and whites showed the smallest gains--only 4 and 5 points, respectively, but still double the gain in the grand mean. The mean for blacks increased by 11 points, the mean for Asian-Americans increased 14 points, and the mean for Puerto Ricans increased 9 points. Changes in the mix of racial/ethnic groups in the test-taking population obscured these gains.

STATUS OF CURRENT INDICATOR SYSTEM

In the light of the salience of achievement measures and the large amount of testing undergone by American students, high-quality indicator data about achievement are surprisingly limited.

A Summary of Data Sources

Currently, debate about the performance of American students focuses on a small number of data sources. On the national level, data from the NAEP are the most salient, with lesser attention focused on college-admissions tests, occasional special studies, and international comparisons of achievement. At lower levels of aggregation, state and local data--in particular, scores on statewide assessments--are often prominent.

The following sections briefly note some of the strengths and weaknesses of major sources of achievement data. International studies raise a number of complex issues discussion of which is beyond the scope of this paper, so they are not discussed here.

The National Assessment of Educational Progress

The National Assessment has become the most salient source of data on the achievement of American students. For that reason, NAEP warrants more extensive discussion here than other databases.

NAEP's role as the leading achievement indicator is well justified. It is the only source of frequent data on the achievement of nationally representative samples of American students (Table 1). It tests students in three grades (currently, 4th, 8th, and 12th) in each biennial testing cycle. It covers a wider range of subject areas than most assessment programs. NAEP's content reflects a broadly-based consensus process. At all stages of the assessment--sampling, construction of the test, and the scaling and analysis of results--the NAEP is carried out with care and sophistication. Its methods, while arcane, are unusually well documented, and its data are always made available for secondary analysis. The entire assessment is subjected to an unusually extensive process of advice and criticism from a wide array of experts, including standing and *ad hoc* panels convened to advise the National Center for Education Statistics or the Educational Testing Service (the prime contractor for NAEP for the past decade) or to carry out Congressionally mandated evaluative studies.

Yet for all its exceptional strengths, NAEP also has important limitations, many of which are poorly understood in the policy world and by the press.

The gold standard: obscuring problems of robustness. Although NAEP is treated as the gold standard in much of the public debate--that is, as the data that must be most accurate--it remains only one test, and consequently, its results are subject to the threats to robustness note above. This is not a criticism of the NAEP *per se*; no matter how carefully a test is built, it will always be subject to the possibility of limited robustness. Important results (either cross-sectional comparisons or trends over time) might be different if the test had a different mix of content, formats, or difficulty levels, or even a different type of administration. For example, black-white differences have sometimes varied across NAEP content areas (larger in measurement and geometry, smaller in low-level numbers and operations and, to a lesser degree, relations and functions; Koretz and Lewis, n.d.). In addition, black students were more likely than whites to skip open-ended items, even after controlling for proficiency (as measured by total scores; Koretz, Lewis, Burstein, and Skewes-Cox, 1992).

Thus the mathematics assessment might show different racial/ethnic differences if the content or format mix were changed.

The problem of unrecognized threats to robustness has been exacerbated by the recent shift to reporting the proportion of students who reach *a priori* standards of performance. This trend, which is national in scope (it is embodied, for example, in new assessment programs in Kentucky and Maryland and was also a key part of California's CLAS assessment until Governor Wilson terminated it this year), takes the form of "achievement levels" in NAEP. The three NAEP achievement levels, called basic, proficient, and advanced, represent judgments about adequate achievement for each tested grade. Because these standards are judgmental, they are likely to vary if different panels of judges are used or if the judges are given a different process to follow. However, a review of a large number of articles about NAEP in the lay press in 1991--the first time achievement levels were used in reporting--found that the judgmental nature of the standards was rarely discussed, and its implications for the robustness of the results were not mentioned (Koretz and Deibert, 1993).

Limits of sampling. The NAEP is to educational achievement what the Current Population Survey is to population characteristics: a general-purpose social survey the design of which represents a compromise among its many potential uses. One place these compromises becomes apparent is in NAEP's sampling design. One compromise was noted above: in the interest of efficiency for estimates at higher levels of aggregation, NAEP does not maintain within-school samples that would be appropriate for school-level analysis (Table 1), and it does not sample on the basis of classrooms at all (even though educational practices vary greatly at the classroom level). Another compromise becomes apparent in the sampling of racial and ethnic minorities. In its assessments designed for cross-sectional comparisons (but not, as explained below, in its trend assessments), NAEP oversamples high-minority schools to obtain sufficiently large minority samples for robust estimates of statistics such as group means. However, even with this oversampling, NAEP obtains very small samples of high-achieving minority students--small enough that they are of no use for analyses that some observers, such as Senta Raizen of NAEP's Technical Review Panel, consider important.[3] To obtain reasonably large samples of high-achieving minority students would require a substantially different design and diversion of resources from other uses.

The compromises inherent in NAEP's sampling design are unavoidable, even if the specific decisions made to date are not. It has become a serious issue recently, however, because of the ever-expanding range of uses to which the policy community and others want to put the NAEP.

Weaknesses of NAEP trend estimates. Over time, NAEP must change, because expectations of what will be taught and learned, and what skills and knowledge are most important to test, change. For example, the policy community currently wants more constructed-response testing and less reliance on the traditional multiple-choice format. This poses a dilemma for the assessment of trends. To alter the test too much runs the risk of confounding changes in the test with trends in performance. On the other hand, leaving the test unchanged renders it increasing irrelevant.

The 1986 NAEP assessment in reading--the second using test design and scaling procedures introduced by the Educational Testing Service when it took over operation of NAEP--produced an implausibly large decline in estimated average reading proficiency dropped sharply at ages 9 and 17. This change, particularly at age 17, was far larger than any of the differences between two assessments since the inception of the reading assessments in 1971 (Beaton and Zwick, 1990). It was concluded that changes in the measurement conditions (i.e., timing and item order) had added an unacceptable amount of error to trend estimates in reading (see Beaton & Zwick, 1990) This lead to the decision to separate NAEP into two assessments (Beaton and Zwick, 1992): a main assessment, which is intended to document what students can do at a particular time and to monitor short-term trends; and a trend assessment, the primary purpose of which is to monitor longer-term trends. The main assessment continued to incorporate changes, while in the trend assessment, every effort has been made to maintain consistency over time.

---

[3] Senta Raizen, personal communication, 1994.

Since then, trend and main assessments have grown quite distinct, but the differences between them--indeed, even the fact that they are not the same--are not widely understood. The trend results are among those given the greatest attention, but they are based on an assessment that is in many ways the poor cousin of the main NAEP assessment. The trend assessment, for example, has substantially sparser sampling of both items and students. It does not oversample high-minority schools. In addition, the trend assessment classifies students by age, while the main assessment (which samples by both grade and age) is reported primarily in terms of grades. The relationship between grade and age, however, has been changing over time. The two assessments also use different methods to delineate racial/ethnic groups: the main assessment relies on students' self-reports unless they are omitted or otherwise unusable, while the trend assessment uses the test administrators' guesses. The disparity between these methods is most extreme in grade 4. For example, in mathematics in 1992, only 40 percent of the fourth-grade students classified as Hispanic by the method used in the main assessment were also classified as Hispanic by the method used in the trend assessment (Barron and Koretz, forthcoming).

The practical consequences of these differences between the two NAEP assessments vary, but in one case they are clearly very important. As a result of the sample design of the trend assessment, NAEP's estimates of relative trends among racial/ethnic groups--on the most important and salient results of the NAEP assessment--have such large errors that in some instances only implausibly large changes could be statistically significant, and the magnitude of changes that reach significance can be estimated only very imprecisely (Table 1; Barron and Koretz, forthcoming).

College Admissions Tests

College-admissions test data, particularly the SAT (formerly the Scholastic Aptitude Test, now the Scholastic Assessment Test), have been used for years as an indicator of trends in student performance, and they still receive substantial attention, even though their inappropriateness for this use has been widely discussed. The SAT was neither designed to be nor validated as a measure of students' mastery of material taught to them during elementary and secondary education. (In contrast, the original American College Testing program college-admissions tests were an adaptation of the Iowa Tests of Educational Development, a relatively difficult achievement test battery for students in grades 9 through 12.) Perhaps even more important, all college-admissions testing suffers from selectivity bias, in that students only take the tests if they chose to and can take it more than once if they wish (Table 1). It is clear that the selectivity of the test-taking population is changing over time, but the nature of the changes are not clear and are difficult to ascertain fully. For example, one study (Beaton, Hilton, and Schrader, 1977) used nationally representative data to estimate the effects of selectivity changes on SAT scores from 1960 through 1972, but I am not aware of any studies of comparable thoroughness addressing selectivity changes in more recent years. The lack of a clear estimate of the impact of selectivity changes severely limits the utility of college-admissions test data as an indicator.

Limitations of State, Local, and Private Data

Most achievement testing in the United States is conducted as part of state or local testing programs. Thus a key question is the adequacy of such data for use in an indicator system.

The usefulness of state and local achievement data in a national system of indicators is questionable. In earlier work (Koretz, 1986, 1987), I made substantial use of state assessment data as a secondary source of data, to confirm or elaborate upon trends apparent in national data. However, as I noted then, the future utility of these data was in doubt even then, because the increased use of tests as accountability tools would likely lead to greater corruption or inflation of scores (Table 1). More recently, Linn and Dunbar (1990) pointed out that many state and local testing programs have shown considerably more favorable trends than has NAEP and suggested that accountability pressures might help explain the disparity.

The utility of state data is also undermined by the rapid rate of innovation in state assessment programs. In response to the widespread view that teaching to multiple-choice tests damaged educational quality, many jurisdictions are shifting rapidly to a greater (even a sole) reliance on various forms of performance assessment. Some of these forms, such as on-demand direct assessments of writing, are hoary and well understood. Others push large-scale assessment into largely uncharted territory. An example is the portfolio assessment programs currently underway in Vermont and Kentucky, in which neither tasks nor administrative conditions are standardized. Another example is group tasks or hybrid group/individual tasks, such as those used in the Maryland assessment, in which part of an assessment task is carried out by a group and the rest is done individually. Research on the validity of these new assessments is only now being undertaken. Thus, whatever the mix of positive and negative effects on instruction of these changes in assessment (and there is evidence that it can have positive effects; see, e.g., Koretz, Stecher, Klein, and McCaffrey, 1994), they will lessen the usefulness of state data in a national indicator system, at least until validation work is completed.

Education Department Longitudinal Surveys

Many of the participants in this conference are familiar with the infrequent, large, nationally representative longitudinal surveys fielded by the federal Department of Education. In the past two decades there have been three such studies: the National Longitudinal Study of the High School Class of 1972 (NLS), High School and Beyond (HSB), and the National Education Longitudinal Survey (NELS). NLS followed a single graduating cohort beginning in their senior year. HSB followed two cohorts, the high school classes of 1980 and 1982, and if followed the younger cohort beginning in tenth grade. NELS began with the class of 1994 when it was in eighth grade and recently completed its third biennial survey.

All of these longitudinal surveys include achievement measures. They provide nationally representative data, adequate sample sizes for racial/ethnic groups, and--unlike NAEP--adequate within-school sampling (Table 1). However, for purposes of providing achievement indicators (as opposed to information on the correlates of achievement and achievement growth), these studies have two important weaknesses. First, the achievement batteries are much smaller than those of NAEP. Second, they are not designed to provide trend data across cohorts, and their achievement test batteries are therefore not necessarily comparable. The Education Department funded a *post hoc* study that equated the NLS and HSB test batteries and analyzed the nature of the changes in performance between the high school classes of 1972 and 1980 (Rock, Eckstrom, Goertz, Hilton, and Pollack, 1985). However, to my knowledge, no attempt has yet been made to equate the HSB and NELS test batteries.

RECOMMENDATIONS FOR A STRENGTHENED INDICATOR SYSTEM

How might the national patchwork of achievement indicators be improved? This question is often interpreted as one about which new indicators would be most useful and which extant indicators could be jettisoned at least cost. In the case of educational achievement, however, strengthening the system of indicators would require more than a revised list of measures. It would require attention to the design and uses of the data systems in which achievement measures are embedded, the design of achievement measures themselves, and the methods of reporting the resulting data. Several recommendations that touch on each of these broad issues follow from the discussion above.

Distinguish the Functions of Achievement Data

If achievement indicators are to maintain their validity, and if funds for additional indicators are to be spent effectively, it will be necessary to distinguish the functions of indicator data clearly from the manifold other functions that achievement test data are expected to perform. First, it is essential that data used as indicators be protected from the potential for corruption that accompanies test-based accountability. Some people in the measurement field, myself included, have expressed concern that the use of NAEP for state comparisons may lead to its corruption. It is not clear whether this fear was warranted; state comparisons have been infrequent and may not have yet become salient enough to warrant efforts to teach to the test, and it is not clear how one would discern whether NAEP, in theory the least vulnerable large-scale assessment in this regard, has been undermined. However, it seems likely that the use of NAEP at the local level would pose substantially greater risks of corruption.

Second, it is necessary to clarify the distinction between data designed for descriptive purposes (even multivariate descriptive purposes) and data intended to support causal inferences. The misconception that NAEP-like data, cross-sectional and inclusive of only a weak set of potential covariates, can support causal inferences is widespread in the policy world. This was evident, for example, in the responses of policymakers and others to the first NAEP state comparisons. It arose again during the early stages of reauthorization of the Elementary and Secondary Education Act, when two senior Education Department employees informed Congressional staff that they wanted a "NAEP-like" assessment of Chapter 1 (now again Title I) students in order to gauge the program's effectiveness. As long as this confusion continues, it will be difficult to make sensible decisions about the allocation of resources in conducting NAEP and other data collection efforts.

Field More Overlapping Measures

The need for multiple measures of achievement is well established in the profession even if widely disregarded in practice. Multiple measures care likely to assess a broader sample of the domains of interest than single measures. But at least as important is the threat to robustness noted above: even relatively similar measures that purport to assess the same constructs will sometimes provide substantially different answers. Moreover, as NAEP's 1986 reading results made clear, even well-designed assessments can occasionally produce unexpected, anomalous results. Multiple measures can provide an indication that a given result is robust enough to be trusted.

Fielding multiple, frequent, nationally representative assessments, however, would require a substantial and probably unrealistic increase in expenditures. A less costly alternative would be to field multiple assessments at varying frequencies. For example, NAEP could be maintained at its current biennial frequency, while other assessments, linked to NAEP but perhaps not entirely overlapping, could be conducted at less frequent intervals. The result would be a richer and more robust set of indicators than would be obtained by the current proposal to field NAEP itself annually rather than biennially.

Mix Formats Carefully

Although the movement toward performance assessment is national in scope, assessment programs differ in the extent to which they place reliance on performance assessments and in the types of formats they employ. NAEP has gradually increased its use of constructed-response items, including longer items that require more extensive use of language (for example, mathematics problems that require explanation of solutions). However, a large portion of the test remains multiple-choice; in most subjects, the constructed-response items are relatively short, and all tasks remain individual rather than group. In contrast, many state programs have made much more dramatic changes. For example, Kentucky's assessment for the last several years comprised four components: multiple-choice items, moderate-length constructed response tasks roughly comparable in length to those in NAEP, large "performance events" that required roughly 3/4 hour each, and portfolios. This year, the state dropped the multiple-choice component entirely. As noted above, some jurisdictions, including Kentucky and Maryland, also include group work in their assessments.

The costs and benefits of various formats may not be the same for indicator systems and for accountability-oriented programs. In the context of state reforms such as those in Kentucky, Maryland, or Vermont, the quality of the resulting measures is only one of several concerns; at least as important is the presumed positive effects on instruction and learning of using such assessments for accountability. Many proponents consider a decrement in reliability to be a reasonable price to pay for those incentives, and some would accept a decrement in some aspects of validity as well.

In the case of indicator data, however, the quality of measurement must be the prime concern. If forms of performance assessment are required to assess certain types of desired outcomes accurately, they should be included in the assessment. However, the breadth, reliability, and validity of the results must be maintained. At least for the time being, these requirements are likely to necessitate a substantial reliance on items that can be answered quickly and scored cheaply, probably including both multiple-choice and short constructed response items. Innovative formats may be included for experimentation and evaluation, but their results should only be used in reporting once they are adequately validated.

Field Complementary, Focused Data Collection

Because NAEP is designed to provide efficient estimates of performance for American students as a whole, it is not able to provide adequate estimates of many important aspects of achievement. And for reasons noted above, other data sources, such as college-admissions tests and state assessment data, provide inadequate supplements. For example, there are no nationally representative data providing an adequate view of trends in the performance of high-achieving high-school students. NAEP is not adequate for this purpose because of limited sampling of high-achieving students (particularly high-achieving minority students) and its dearth of test items appropriate for students at that level. College-admissions test data are inadequate for this purpose because of selectivity bias.

Therefore, to provide a stronger system of achievement indicators, large-scale broad-purpose surveys such as NAEP should be complemented with less frequent data-collection efforts focused on populations or topics that require different sampling of students or tasks. Both the content of these studies and their frequency are matters of judgment and disagreement. For example, in recent years, one very large and expensive supplement was added to NAEP: the Trial State Assessment (TSA), which provides achievement estimates for states and for a few subgroups within them. On the one hand, critics have argued that TSA may be a poor use of the large amount of money it requires, because it can't support the causal inferences that many of its proponents want, will confirm many things that are widely known already (e.g., that states such as Minnesota and Iowa outscore states such as Louisiana and Mississippi), and lack the ability to clarify whether the few surprising findings are really robust (Koretz, 1991). On the other hand, TSA can clarify where problems of low achievement are most severe, even if they can't explain those findings. Given NAEP's credibility, such findings might be very useful even if they are unsurprising to experts. Second, TSA could serve as an audit test, signalling when trends on state-administered tests are grossly inflated.[4]

More value might be learned by putting the additional resources into targeted studies of important populations and topics that neither NAEP nor other current databases can address. The possibilities are numerous:

Studies of special populations of students or aggregates. Complementary studies of specific populations poorly sampled by broad-purpose surveys such as NAEP and NELS could be valuable to policymakers and the public. Among the groups that might be appropriate focuses of such studies are high-achieving students, immigrant children (and children of immigrants), children with limited proficiency in English, handicapped students, and ethnic groups that are too small for routine oversampling, such as various Hispanic groups and immigrant Asian groups. NAEP-like studies of at-risk populations could also be very valuable, albeit not for the program-evaluation function proposed by some advocates.

Complementary studies might also be used to provide a broader range of information about certain groups of students who are already adequately sampled for certain statistics. African American students provide a good example. Although NAEP currently provides a good cross--sectional estimate of the black mean, it offers only a fairly error-prone estimate of trends in that mean over time and virtually no useful information about high-achieving blacks (because there are too few in the sample). Periodic larger samples of African American students could ameliorate some of these limitations.

Periodic special studies might also be used to provide reliable data at levels of aggregation that currently are not well supported by NAEP or other databases--for example, to provide robust estimates of achievement trends in different types of schools or to investigate changes in the instructional resources provided to students in different types of classes and schools.

Achievement supplements to longitudinal surveys. NELS, HSB, and NLS all include achievement measures, but NELS did not begin until 8th grade, and HSB did not begin until 10th. In addition, these surveys do not include the detailed measurement of income, program participation, and other important social constructs that are measured by longitudinal surveys like PSID and SIPP. An achievement supplement to one or more such surveys (which could use an adaptation of the NAEP item bank) could provide valuable information--for example, differences in achievement (rather than just attainment) between children who are poor long-term and short-term.

---

[4] However, to serve this function, state NAEP would have to be conducted differently. Specifically, limited resources would have to be used assess the same subject areas reasonably frequently.

Studies of different achievement constructs. Complementary, less frequent studies could also be used to provide measures of a wider array of achievement constructs, either in response to sampling of special populations or for more general interest. The infrequent national surveys of literacy in the young adult population provide a good example of one of the few such efforts already in place. It could be very useful, for example, to put in place a periodic survey of higher levels of high-school mathematics and science--using items more difficult than most now in NAEP while avoiding the selectivity bias that plagues data such as the College Board Achievement Tests or ACT.

Experiment with Multiple Metrics

Because different metrics often provide different views of patterns of achievement, it will often be important to present key findings about achievement using several different measures. For example, there are several reasons to complement mean or median differences among groups with "representational indicators," such as the proportion of students in each group falling into high quantiles or the proportion reaching high *a priori* standards. One reason is the fact that lay audiences will rarely understand the implications of a simple difference in central tendency for the proportion of students reaching a given threshold; another is that the policy world currently places great emphasis on the proportion of students reaching high standards. (There are also reasons not to report only the proportion of students reaching standards, including lesser precision of estimates and, in the case of unreliable measures, likely bias; see Koretz, Stecher, Klein, and McCaffrey, 1994.)

However, recent experience (e.g., Koretz and Deibert, 1993) has shown that some current approaches to reporting are not effective with lay audiences. Some efforts are now underway to gain more understanding of the impact of alternative presentations on lay understanding of assessment results, but knowledge is still very limited. Accordingly, indicator development in this area should be accompanied by an active program of research and evaluation.

Present Simple Statements of Confidence or Robustness

Because lay audiences comprise many of the key users of indicator data, the problem of finding a way of expressing degrees of confidence in terms they can understand affects all of the indicators under discussion in this conference. It is clear that simple statements of statistical significance are often found incomprehensible, but it remains less clear what alternative presentations might be more effective. In the case of assessment indicators, this problem is compounded by the relative importance of other, non-sampling threats to robustness, such as simple measurement error or potentially systematic differences attributable to the specifics of test construction. These non-sampling sources of error are rarely presented in reports of achievement indicators but should be, even when they cannot be quantified.

CONCLUSION

Many observers have commented on the enormous weight cognitive tests are given in American popular and policy debate, and it is clear that past 20 years have witnessed a great increase in the prominence and importance of test scores. In addition, some observers have commented on the large number of tests that American students take. Yet despite the frequency and salience of testing in this country, the range of data well suited to use in an indicator system is surprisingly limited. Any number of steps to expand the stock of achievement indicators are practical, but improvement will depend on a careful separation of the requirements of indicators from other uses of tests and agreement on the relative value of additional data of various sorts.

REFERENCES

Airasian, P. (1987). State mandated testing and educational reform: Context and consequences. *American Journal of Education, 95*, 393-412.

Barron, S. I., and Koretz, D. M. (forthcoming). An evaluation of the robustness of the NAEP trend lines for racial/ethnic subgroups. Manuscript submitted for publication.

Beaton, A. E., Hilton, T. L. and Schrader, W. B. (1977). *Changes in the Verbal Abilities of High School Seniors, College Entrants, and SAT Candidates between 1960 and 1972.* New York: College Entrance Examination Board.

Beaton, A. E. & Zwick, R. (1990). *The effect of changes in the national assessment: Disentangling the NAEP 1985-86 reading anomaly* (Report No. 17-TR-21). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Beaton, A. E. & Zwick, R. (1992). Overview of the National Assessment of Educational Progress. *Journal of Educational Statistics*, 17, 95-109.

Bracey, G. (1991). Why can't they be like we were? *Phi Delta Kappan, 73* (2), October, pp. 104-117.

Burket, G. R. (1984. Response to Hoover. *Educational Measurement: Issues and Practice, 3*(4), 15-16.

Clemans, W. V. (1993). Item Response Theory, vertical scaling, and something's awry in the state of test mark. *Educational Assessment, 1*(4), 329-347.

Dunbar, S., Koretz, D., and Hoover, H.D., (1991). Quality Control in the Development and Use of Performance Assessment. *Applied Measurement in Education, 4* (4), 289-303.

Goslin, D. A. (1963). *The Search for Ability: Standardized Testing in Social Perspective*. NY: Russell Sage.

Goslin, D., Epstein, R. R., and Hallock, B. A. (1965). *The Use of Standardized Tests in Elementary Schools.* NY: Russell Sage.

Haney, W. (1981). Validity, Vaudeville, and Values: A short history of social concerns over standardized testing. *American Psychologist, 36* (10), 1021-1034.

Haney, W. (1984). Testing reasoning and reasoning about testing. *Review of Educational Research, 54* (4), 597-654.

Hedges, L. V. Personal communication, September 30, 1994.

Hoover, H. D. (1984a). The most appropriate scores for measuring development in the elementary schools: GE's. *Educational Measurement: Issues and Practice, 3*(4), 8-14.

Hoover, H. D. (1984b). Rejoinder to Burket. *Educational Measurement: Issues and Practice, 3*(4), 16-18.

Hoover, H. D. (1988). Growth expectations for low-achieving students: A reply to Yen. *Educational Measurement: Issues and Practice, 7*(4), 21-23.

Jaeger, R. M. (1982). The final hurdle: Minimum competency achievement testing. In G. R. Austin and H. Garber (Eds.), *The Rise and Fall of National Test Scores*. New York: Academic Press.

Jaeger, R. M. (1992). "World class" standards, choice, and privatization: Weak measurement serving presumptive policy. Vice-Presidential Address (Division D), annual meeting of the American Educational Research Association, San Francisco, April 20.

Koretz, D. M. (1986). *Trends In Educational Achievement*. Washington, D. C.: Congressional Budget Office, U. S. Congress.

Koretz, D. M. (1987). *Educational Achievement: Explanations and Implications of Recent Trends*. Washington, D.C.: Congressional Budget Office, August, 1987.

Koretz, D. M. (1990). *Trends in the Postsecondary Enrollment of Minorities*. Santa Monica: RAND (R-3958-FF).

Koretz, D. M. (1991). State comparisons using NAEP: Large Costs, Disappointing benefits. *Educational Researcher, 20* (3), 19-21.

Koretz, D. M. (1992a). State and National Assessment. In M.C. Alkin (ed.) *Encyclopedia of Educational Research*, Sixth Edition, Washington, D.C.: American Educational Research Association.

Koretz, D. M. (1992b). *Validating and Evaluating Indicators of Mathematics and Science Education*, N-2900-NSF, Santa Monica: The RAND Corporation.

Koretz, D. M. (1992c). What Happened to Test Scores, and Why?" *Educational Measurement: Issues and Practice, 11* (4), 7-11.

Koretz, D. M., and Deibert, E. (1993). *Interpretations of National Assessment of Educational Progress (NAEP) Anchor Points and Achievement Levels by the Print Media in 1991*. Santa Monica: The RAND Corporation (MR-385-NCES).

Koretz, D., and Lewis, E. *Indicators of Achievement in Mathematics and Science*. Santa Monica, CA: RAND, unpublished working paper.

Koretz, D. M., Lewis, E., Burstein, L., and Skewes-Cox, T. (1992). *Omitted and Not-Reached Items in Mathematics in the 1990 National Assessment of Educational Progress*. RAND and UCLA, Center for Research on Evaluation, Standards, and Student Testing, Technical Report to the National Center on Education Statistics.

Koretz, D. M., Linn, R. L., Dunbar, S. B., and Shepard, L. A. (1991). The Effects of High-Stakes Testing: Preliminary Evidence About Generalization Across Conventional Tests. In R.L. Linn (chair), *The Effects of High Stakes Testing*, symposium presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago, April.

Koretz, D. M., Stecher, B., Klein, S., and McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice, 13* (3), 5-16.

Linn, R. L, and Dunbar, S. B. (1990). The Nation's report card goes home: Good news and bad about trends in achievement. *Phi Delta Kappan, 72* (2), October, 127-133.

Linn, R. L., Kiplinger, V. L., Chapman, C. W., and LeMahieu, P. G. (1992). Cross-state comparability of judgments of student writing: Results from the New Standards Project. *Applied Measurement in Education, 5*(2), 89-110.

Linn, R. L., Shepard, L., & Hartka, E. (1992). The relative standing of states in the 1990 Trial State Assessment: The influence of choice of content, statistics, and subpopulation breakdowns. In National Academy of Education, *Assessing Student Achievement in The States: Background Studies*. Stanford, CA: Stanford University, National Academy of Education.

Mullis, I.V.S., Dossey, J.A., Owen, E. H, and Phillips, G. W. (1991). *The State of Mathematics Achievement: NAEP's Assessment of the Nation and the Trial Assessment of the States*. Washington, D. C.: National Center for Education Statistics (21-ST-04).

Mullis, I.V.S., Dossey, J.A., Owen, E. H, and Phillips, G. W. (1993). *NAEP 1992 Mathematics Report Card for the Nation and the States*. Washington, D. C.: National Center for Education Statistics (23-ST02).

Murnane, R., and Raizen, S. (1988). *Improving the Quality of Indicators of Science and Mathematics Education in Grades K-12*. Washington, D. C.: National Academy Press.

National Center for Education Statistics (1988). *Dropout Rates in the United States*. Washington, D.C.: U. S. Department of Education (NCES 89-609).

National Commission on Excellence in Education (1983). *A Nation at Risk*. Washington, D. C: U. S. Department of Education.

Pipho, C. (1985). Tracking the reforms, part 5: Testing--can it measure the success of the reform movement? *Education Week, 4* (35), 19.

Raizen, S., and Jones, L. V. (1985). *Indicators of Precollege Education in Science and Mathematics*. Washington, D. C.: National Academy Press.

Resnick, D. (1982). History of educational testing. In A. K. Wigdor and W. R. Garner (Eds.), *Ability Testing: Uses, Consequences, And Controversies, Part I*. Washington, D. C.: National Academy Press, pp. 173-194.

Rock, D. A., Eckstrom, R. B., Goertz, M. E., Hilton, T. L., and Pollack, J. (1985). *Factors Associated with Decline of Test Scores of High School Seniors, 1972 to 1980*. Washington, D.C.: Center for Statistics, U.S. Department of Education.

Roeber, E. (1988). A history of large-scale testing activities at the state level. Paper presented at the Indiana Governor's Symposium on ISTEP, Madison, Indiana, February 10.

Shavelson, R. J., Baxter, G. P., and Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement, 30* (3), 215-232.

Shavelson, R. J., McDonnell, L., Oakes, J., and Carey, N. (1987). *Indicator Systems for Monitoring Mathematics and Science Education*. Santa Monica: RAND (R-3570-NSF).

Shepard, L. A., and Dougherty, K. C. (1991). Effects of High-Stakes Testing on Instruction. In R.L. Linn (chair), *The Effects of High Stakes Testing*, symposium presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago, April.

Spencer, B. (1983). On interpreting test scores as social indicators: Statistical considerations. *Journal of Educational Measurement, 20* (4), 317-333.

Wainer, H., and Thissen, D. (1993). Combining multiple-choice and constructed-response test scores: Toward a marxist theory of test construction. *Applied Measurement in Education, 6*(2), 103-118.

Yen, W. M. (1988). Normative growth expectations must be realistic: A response to Phillips and Clarizio. *Educational Measurement: Issues and Practice, 7*(4), 16-17.

Richard J. Murnane November 17, 1994

Summary of four papers on education indicators

I will discuss the papers in the order of the chronological ages of the children to which they pertain, beginning with the Phillips-Love paper, moving to Koretz, then to Hauser, and then to Kane. At the end I will say a few words about the respects in which the four papers reflect different stages in what might be called the eternal cycle of indicator development.

Deborah Phillips and John Love

The paper by Deborah Phillips and John Love suggests indicators in three areas: school readiness, child care, and the first years of schooling. They suggest the need for many indicators in each of these three areas. Table 1, which is essentially their Table 1, lists the eight areas for which they suggest indicators of school readiness. The first column lists the conceptual areas. The second column lists currently available data sources. Note the great many blanks in column 2. The reason is that developing indicators of school readiness is a relatively recent undertaking. A great deal of work is needed in developing instruments and figuring out the best way to field them on a regular basis. Also, the source of data for most of the proposed indicators of school readiness is the National Household Education Survey (NHES), a telephone survey of households with 3-7 year olds.. I understand from Brett Brown's paper that it is not yet clear how often the school readiness module will be included in this survey that is administered every other year. Thus the question of how often indicators of school readiness are produced will depend critically on decisions concerning the composition of the NHES.

The third column lists surveys that may be used in the future to collect information needed for the proposed indicators. The ECLS is the Early Childhood Longitudinal Survey, which is scheduled to start in 1998. Relying on data from longitudinal surveys for indicators mean that the quality is likely to be high, but the indicators will only be updated rarely, since expensive longitudinal surveys are started only infrequently.

The indicators in column 1 that are underlined in bold are the conceptual areas that the authors feel are the most important to develop indicators for. They include: exposure to reading at home, approaches or attitudes toward learning (which include curiosity about tasks, persistence, imagination,,,), and Access to Instruction in Native Language.

Table 2 of the Phillips-Love paper provides their recommendations for indicators of Child Care. Again notice the blank spots in column 2, indicating that sources of indicators for many of the concepts do not currently exist. There are not as many blank spots in column 2 of the child care indicators table as there were in column 2 of the school readiness indicators table, suggesting that child care indicators are more ready to go than school readiness indicators. The underlinings illustrate that the authors view as priority areas for indicators, the stability of childcare arrangements, the proportion of eligible children in early intervention programs, and child care costs as a function of family income.

I wondered about the extent to which trends over time in the proportion of eligible children served in early intervention programs would be influenced by changes in the definition of eligibility.

Table 3 lists the authors' suggestions for indicators of early schooling. Here the priority areas for indicators are: achievement, progress in school and bilingualism.

One question I had concerned the concept, bilingualism. It makes sense to me that in the increasing global economy, all children should learn to speak at least two languages. Is this what the authors mean when they suggest an indicator measuring "exposure to bilingual education." Is it necessary to distinguish this from the varying needs in different parts of the country to teach English to children who come to school speaking a different language? I ask because it seems important to be clear on what an indicator is supposed to mean, and whether a higher value for the indicator means that things have gotten better or they have gotten worse. A rising value for this indicator could mean that American schools are doing a better job of exposing children to different cultures and languages or that they are needing to invest more in teaching English to an increasingly varied student clientele. To know what is happening (and indicators are supposed to facilitate this), it would be important to be able to distinguish between the two interpretations.

Dan Koretz

While the Phillips-Love paper proposes a great many new indicators and seems optimistic in tone about what can be learned from new indicators, Dan's paper has very few suggestions for new indicators and most of the space is devoted to explaining the very steep tradeoffs involved in the design of achievement testing programs. Implicit in his paper is the recognition that, with 25 years of information from the National Assessment of Educational Progress (NAEP), the United States has made enormous progress in measuring the achievement of school-aged children. There are reasons to worry about challenges to the integrity of the NAEP.

Dan's paper provides six recommendations for a strengthened indicator system. These are summarized in Table 4.

The first is to distinguish among the uses of achievement data. Dan argues compellingly that it is a mistake to try to use the same assessment program for multiple purposes. One reason is that the design appropriate for one use will be very different from the appropriate design for another use. For example, matrix sampling is extremely useful in a testing program such as the NAEP designed to provide information for indicators, but it is not good for a program designed to be part of an accountability system. Also, the coaching to the test that follows when scores are used in an accountability program makes the scores inflated estimates of the extent to which students have mastered a particular domain of knowledge or skill.

Dan's second recommendation is to Field More Overlapping Measures. Dan's work provides many examples of why it is a mistake to make judgments about trends in the achievement of the nation's children from any one test score series. There is no question that Dan's call for multiple measures is correct. But it is important to recognize that multiple measures inevitably will reveal puzzles that raise questions about the quality of the indicators.

The third recommendation is to mix formats carefully. Dan suggests that while performance assessments (often called authentic assessments) have strengths that may be important in designing accountability programs -- most important, coaching may have desirable affects on instruction, while teaching to multiple choice assessments may not -- performance assessments have severe limitations as the basis for indicators. In particular, there are many questions about reliability and validity of the performance assessments that states have begun to use. Thus, Dan's advice is that achievement indicators are **(in his words)** "likely to require a substantial reliance on items that can be answered quickly and scored cheaply, probably including both multiple-choice and short constructed response items.

Dan's fourth recommendation is to field complementary focused studies. In other words, fund targeted studies of important populations and topics that neither NAEP nor other current databases can address. Examples include studies of immigrant children or high achieving minority children. The implicit message is that it will never be possible to design NAEP in a way that provides detailed trend information on the performance of every group of particular interest. Instead of trying to do this, better to fund supplemental studies. Tom Kane has very similar advice.

The fifth recommendation is to experiment with multiple metrics. The key point is to find ways of describing trends that lay audiences -- and perhaps, especially the press -- find interesting and informative. But it is important to point out that different indicators constructed from the same achievement data can tell different stories. For example, Dan points out that ranking of states by the median scores on the NAEP leads to a different ranking than the ranking produced from the 75th percentiles.

Dan's final recommendation is to figure out ways to display information about confidence intervals so that the audience for indicators has a sense of whether differences in scores over time or among states are meaningful.

Bob Hauser: Indicators of High School Dropout

Bob Hauser explains that recent changes in the educational attainment questions included in the CPS and in the Census complicate the task of constructing reliable trends in educational attainment and dropout rates. Tom Kane also expressed concern about the changes in the CPS educational attainment question.

Why did the Census Bureau change the educational attainment question in the CPS and the Census? As Bob Kominski and Paul Siegel explain in a recent article in the Monthly Labor Review, the reasons include the following:

> With the old questions, years of schooling completed tended to be misclassified into degree status.
>
> With the old questions, it was not possible to identify specific degrees.
>
> The old questions led to uncertainty in the classification of high school graduates.

The changes in the CPS educational attainment questions are an example of a classic dilemma in indicator construction. What do you do when changes in the way the world works render increasingly problematic responses to questions that have been asked for a long time? Keeping them the same means that the data are increasingly poor descriptors of what is happening. Changing the questions means an abrupt break in trend data.

Bob Hauser's concerns about the new CPS educational attainment questions include the following (summarized in the table):

1. collapse of several grade levels below high school has made it impossible to follow age-grade progressions at younger ages.

(The new question collapses grades 1-4 into one category, grades 5-6 into a second, and grades 7-8 into a third.)

2. failure to distinguish grades attended from grades completed has eliminated the ability to examine a key educational transition -- between college entry and completion of first year of college. With the new questions a great many people who dropped out of college during their first year are now classified as having "completed some college."

3. new questions do not distinguish between entry into 12th grade and completion. (see pp.15-16)

4. collapse of grades 13-15 into "some college no degree" has created a large and extremely heterogeneous category. some have more education that the two years required for an AA degree; some have less. The category includes people classified under the old system as obtaining no college education -- they did not complete first year of college.

5. p.17: "the new educational classification fails to distinguish between individuals who completed 12 years of school from those who achieved high school equivalency. Bob suggested putting GED holders in this category.

One reason it is important to distinguish GED holders from conventional high school grads is that the earnings of high school graduates serve as the baseline for indicators of the payoff to college that Tom Kane advocates. As Jim Heckman has pointed out, including a growing number of GED-holders in the population of "high school graduates" whose earnings form the basis for computing the payoff to college has the effect of reducing estimates of the earnings of high school graduates and inflating the estimated returns to post-secondary education.

Bob Hauser concludes that the new CPS education questions should not be used as a model in other surveys.

Bob concludes by raising questions about the indicator of high school completion rate by state included in the "Kids Count" volumes. The measure is obtained by dividing the number of public high school graduates in the reference year by the public ninth grade enrollment four years earlier, with some adjustment for inter-state migration. Bob points out that the correlations between this measure and the high school completion rates (for 25-29 year-olds and 20-24 year-olds) calculated from Census data are disturbingly low (.69 and .78). Also, the Kids Count indicator shows drop out rates increasing while CPS-based indicators show drop out rates falling. The explanation may have to do with the growing number of people acquiring the GED credential each year -- approximately 600,000 people currently.

The median age of GED-recipients is in the early 20s. These people are counted as dropouts in the Kids Count indicator and as graduates in the CPS-based indicator of the schooling attainment of 24-year-olds.

Tom Kane

Tom focused his paper on indicators of ACCESS (Who is enrolled...), COST (How much...) and payoffs to different types of post-secondary education. Notice that this with this third topic, payoff, Tom is directly confronting the causality question, thereby creating new and difficult challenges for indicator development.

One disturbing pattern, illustrated in the next figure is that the college enrollment rate for Black high school graduates is lower in every year than the corresponding enrollment rate for white high school graduates. The likely explanation is that Black students live in lower income families. As Tom explains, this hypothesis cannot be investigated with the CPS because many students of college going age have left their parents' home and set up new households. Young people not going to college are particularly likely to do this. This forms the motivation for Tom's first indicator suggestion.

On Access:
Tom's recommendation is to: Collect parental education and occupation information for young adults (ages 16-24) on the Current Population Survey.

The reason is that it is not possible with currently available data from the CPS to construct trends in college attendance rates by people with different socioeconomic backgrounds, defined by the education and occupation of parents of 18-24 year olds. The requisite data is available for young adults living in their parent households, but not for those who leave home to form their own households. Tom suggests asking these young people who form their own households about the education and occupation of parents. What he really wants to know is income, but does not suggest asking for this because the answers would be unreliable.

If Tom's recommendation about new questions for the CPS were followed, it would be possible to use the CPS to compare college enrollment rates for Black, White and Hispanic youth from similar socioeconomic backgrounds. As Tom points out, there is some evidence on this from the Dept. of Education longitudinal studies, but the cohorts are infrequent, and the lags in acquiring new information are therefore long. Tom's suggested CPS questions would also make it possible to track high school completion rates by SES, as Bob Hauser suggested.

On Cost:
Tom points out that the dollar figures for tuition, room and board printed in college catalogs are not meaningful figures for college costs because there is now a great deal of means-tested student aid. His recommendation is to:

> Develop a small number of student profiles, specifying family income and savings levels, and interview the state financial aid offices directly to learn about available state grants each year.

The idea would be to track over time for each state the cost net of financial aid of going to a particular kind of college for a young person with a particular family income and asset profile.

Tom also suggests reporting on a regular basis:

> the earnings foregone by students attending college.

The reasoning is that foregone earnings are a large part of college costs, and changes in foregone earnings are likely to affect college attendance rates, just as changes in net tuition levels do.

Measuring the Payoffs to College
The CPS does provide the basis for constructing trends in the relative earnings of four-year college graduates and high school graduates. For both males and females, the college/high school earnings differential has widened markedly since 1979. But parents and high school seniors often want the answers to questions more specific than what has happened to the college/high school earnings differential. For example, they ask:

Does it matter whether a person goes to a two or four year school?

What about the value of completing a post-secondary vocational school program?

Tom explains that the available data are less suited to address these questions about the payoffs to alternative types of post-secondary education. While the CPS asks current students about whether they are attending a two year or four year college or a vocational school, it does not provide this information for past students. Tom suggests:

> Experimenting with questions to distinguish prior attendance at 2-year, 4-year, and vocational schools as a supplement to the CPS educational attainment question.

Like Bob Hauser, Tom endorses some aspects of the Census Bureau's change in educational questions, but laments the loss of information on number of years of schooling completed. His work shows that years of college completed is a strong predictor of earnings. So the heterogeneous residual category, "some college," is problematic in predicting earnings.

Finally, Tom points out that very little is known about the returns to attendance at post-secondary vocational schools, in large part because many do not respond to requests for transcripts of those students included in the Dept. of Education longitudinal surveys who attend such schools.

I share Tom's concern since my work has shown that this is the type of post-secondary training that high school dropouts who acquire a GED are the most likely to acquire.

Tom suggestion is to conduct a targeted longitudinal study of youth attending urban high schools, perhaps taking relatively dense samples from a relatively few schools. These students would be especially likely to attend vocational schools, and the greater density of student attendees might help in getting compliance with requests for transcripts. Bob Hauser makes a similar suggestion: increase the oversampling of urban minority youth in longitudinal surveys.

The never-ending cycle of indicator development

In conclusion, I would like to make a few comments about the never-ending cycle of indicator development. The four papers illustrate four aspects of this never-ending cycle -- a cycle that pertains to all of the types of indicators discussed at this meeting, not just education indicators.

The paper by Deborah Phillips and John Love reflects the optimism associated with a new wave of indicator development. There are important things to measure, and if we would devote resources to the task, we could markedly improve our understanding of trends in children's well being.

Dan Koretz's paper reflects the aspect of the cycle where the reality of resource scarcity is dominant, and there is great pressure to use data collection efforts for multiple purposes. There are always costs in doing this, and Dan argues that these costs are extraordinarily high when the proposed dual uses are indicators of performance and measures of accountability.

Bob Hauser's paper reflects the wrenching dilemma of whether to stick with questions that become flawed over time, or whether to change the questions, creating interruptions in long time series of indicators.

Tom's paper reflects the insatiable appetite for data that will answer more refined questions. It is important not only to know about college enrollment rates by race and ethnicity; we also should know about them by socioeconomic status. It is not enough to know about the payoff to an AA degree and a BA; it is also important to learn about the payoff to post-secondary schools. Under the right budgetary circumstances, this can lead back to the optimism stage of the cycle where new data collection efforts can dramatically improve indicators of well-being.

These stages, optimism, pressure for double-duty, the dilemma of whether to change questions, and new appetites for better indicators, are present in every indicator activity. The cycle never ends because improved indicators reveal new puzzles. It is almost always the case that the new puzzles cannot be well understood with available indicators. This calls into question the quality of existing indicators and increases the demand for better indicators.

It is important to keep this Iron Law of Indicators in mind in evaluating the quality of available indicators. The inability to answer new questions does not mean that available indicators were not worth the investment in their development. Progress in indicator development has led to an dramatic increase in the sophistication of the questions that we ask of the data and this is often a significant accomplishment. This historical perspective is critical in judging the value of indicators that we have, and ones we are thinking of developing.

Table 1

Indicators of School Readiness

Indicator	Current Sources	Future Prospects
Exposure to Reading at Home	NHES:93	NHES:95/96 SIPP Child Module NLSY-MC ECLS
Exposure to Pre-Numeracy Experiences		NHES:95/96 ECLS
Approaches to Learning		ECLS Love et al., 1994
Emergent Literacy and Numeracy Development	NHES:93 Prospects Study	NHES:95/96 SIPP Child Module ECLS Love et al., 1994 State/local level data
Proportion of Kindergartners "Unready" for Kindergarten	NHES:93	NHES:95/96 SIPP Child Module State/local level data
Parental Attitudes/ Expectations		NHES:96/96 ECLS
Access to Instruction in Native Language		NHES:95 OECD Schools and Staffing Survey
Access to Technology		X

Table 2

Child Care Indicators

Indicator	Current Sources	Future Prospects
Quality of Care		ECLS State regulatory data
Stability of Care	SIPP Child Care Module	NHES:95 SIPP Child Module
Proportion of Eligible Children in Early Intervention Programs		Survey of Program Dynamics State/local level data
Proportion of children in Latchkey Situations	SIPP Child Care Module	
Child Care Costs: Family Income	SIPP Child Care Module	NHES:95
Parent Choice		NHES:95/96 SIPP Child Module
Access to Providers who Speak Home Language		X

Table 3

Indicators of Early Schooling

Indicator	Current Sources	Future Prospects
Achievement	NAEP	ECLS NEGP initiatives State/local level data
Progress in School	NHES:93 Profiles Study	SIPP Child Module NHES:95/96 ECLS
Engagement in School	NHES:93	ECLS
Parental Involvement/ Participation		NHES:96 ECLS
Bilingualism		NHES:95 OECD ECLS

Improving Indicators of Student Achievement

1. Distinguish among the Uses of Achievement Data

   Don't try to use one data collection effort for both indicators and accountability

2. Field More Overlapping Measures

   Multiple Measures are needed to distinguish disturbing trends from puzzling idiosyncracies

3. Mix Formats

   Only use measures shown to be reliable and valid.

   For indicators, multiple choice items and short constructed response items are preferable to "authentic assessments of performance."

4. Field Complementary Focused Studies

   No broad-based indicator system can provide at reasonable detailed reliable information about groups of special interest. To learn more about the achievement of special groups, such as immigrant children or high achieving children, conduct focused studies.

5. Experiment with Multiple Metrics

   It is not obvious which presentation of indicator data will be most meaningful to lay audiences and the media.

6. Present Simple Statements of Confidence or Robustness

   Readers need a way to judge which differences across states or changes over time are worth paying attention to.

Reasons for Changing the Educational Attainment Items in the CPS

With the old questions:

Years of schooling completed tended to be misclassified into degree status.

Many people who report completing four or more years of college do not have a Bachelor's Degree.

It was not possible to identify specific degrees.

No way to determine who had earned an Associate's degree.

There was uncertainty in the classification of high school graduates.

Many people who dropped out of high school after, say, grade 10 and later earned a GED reported that they had completed 10 years of education, rather than that they were high school graduates.

Many people who completed 12 years of schooling, but did not earn a high school diploma (because they did not pass an exit exam), were counted as being high school graduates.

Figure 10. The 1990-Basis CPS Educational Attainment Classification

---

What is the highest level of school ... has completed or the highest degree ... has received?

Code	Level of Schooling Completed
31	Less than first grade
32	1st, 2nd, 3rd, or 4th grade
33	5th or 6th grade
34	7th or 8th grade
35	9th grade
36	10th grade
37	11th grade
38	12th grade NO DIPLOMA
39	HIGH SCHOOL GRADUATE - high school diploma or the equivalent (For example, GED)
40	Some college but no degree
41	Associate degree in college - Occupational/vocational program
42	Associate degree in college - Academic program
43	Bachelor's degree (For example: BA, AB, BS)
44	Master's degree (For example: MA, MS, MEng, MEd, MSW, MBA)
45	Professional School Degree (For example: MD, DDS, DVM, LLB, JD)
46	Doctorate degree (For example: PhD, EdD)

Concerns About the New CPS Educational Attainment Questions

1. **Cannot follow age-grade progressions at younger ages**

   (The new question collapses grades 1-4 into one category, grades 5-6 into a second, and grades 7-8 into a third.)

   Cannot accurately examine schooling attainments of groups like immigrants who have low schooling levels.

2. **Cannot examine transition between college entry and completion of first year of college**

   With the new questions a great many people who dropped out of college during their first year are now classified as having "completed some college."

3. **Cannot distinguish between entry into 12th grade and completion**

4. **Collapse of grades 13-15 into "some college no degree" has created a large and extremely heterogeneous category**

   This category includes people with more education than the two years required for an AA degree, and people with less.

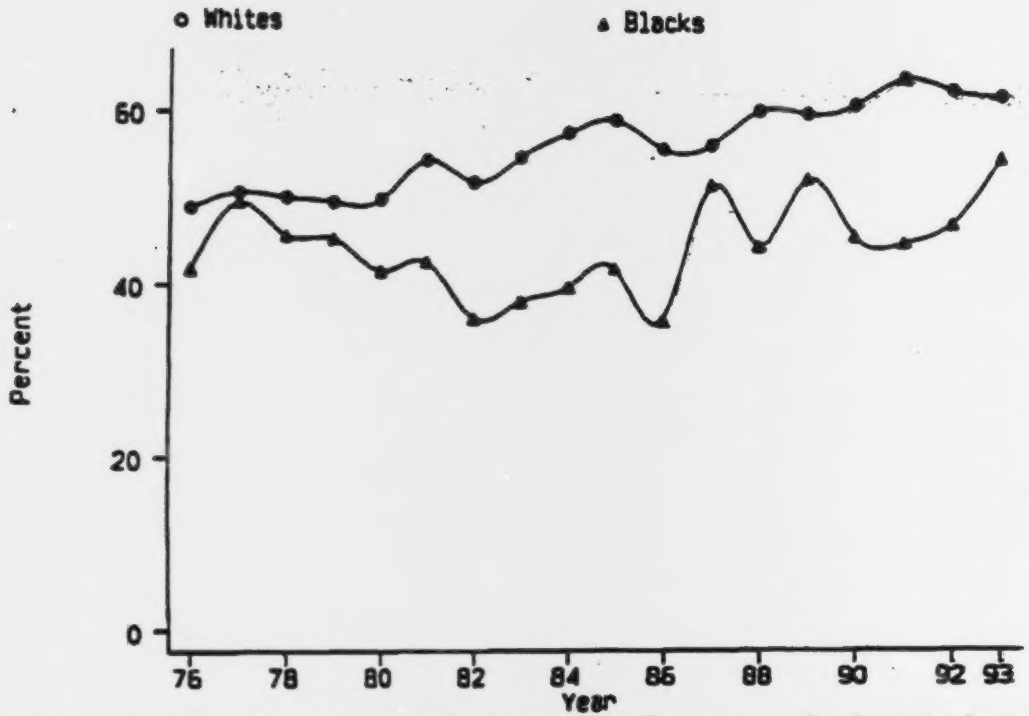
   This category includes people classified under the old system as obtaining no college education -- they did not complete first year of college.

5. **The new educational classification fails to distinguish between individuals who completed 12 years of school from those who achieved high school equivalency**

   Proportion of new "high school graduates" who are GED-holders has grown from 2% in 1954 to 14% in 1987.

   Subsequent earnings of male GED-holders are closer to earnings of dropouts than to earnings of conventional high school graduates.

   > Hauser suggests that GED-holders should be grouped with people completing 12 years of schooling, but having no high school diploma.



College Enroll. of 16-24 Year-old Recent H.S. Graduates

Improving Indicators of College Access, Cost, and Payoff

Access:

Collect parental education and occupation information for young adults (ages 16-24) on the Current Population Survey

Cost:

Develop a small number of student profiles, specifying family income and savings levels, and interview the state financial aid offices directly to learn about available state grants each year

Report on a regular basis the earnings foregone by students attending college

Payoff:

Experimenting with questions to distinguish prior attendance at 2-year, 4-year, and vocational schools as a supplement to the CPS educational attainment question

Conduct a targeted longitudinal study of youth attending urban high schools, taking relatively dense samples from a relatively few schools

Stages in the Never-Ending Cycle of Indicator Development

Optimism:

With sufficient resources, indicators could be much better.

Double Duty?

Resources for data-collection efforts are scarce. Data should serve multiple purposes.

The Wrenching Dilemma

Should we stay with questions that have become flawed, or change questions and suffer interruptions in long-term trends?

New Appetites

Indicators must provide more detailed information.

ECONOMIC SECURITY

Income, Employment, and the Support of Children

Susan E. Mayer
Harris Graduate School of Public Policy Studies
University of Chicago

Paper Prepared for the
Conference on Indicators of Children's Well-Being
Rockville, Maryland
November 17-18, 1994

The section of this paper on income and material well-being is part of a long-term collaboration with Christopher Jencks. Many of the ideas and much of the analysis are the result of this joint work. However, errors that may appear in this paper are mine alone.

This work would not have been possible without the programming assistance of David Knutson, Judith Levine, David Rhodes, Tim Veenstra, and Scott Winship. We are also indebted to John Sabelhaus for providing us with his extracts from the Consumer Expenditure Surveys and Rob Mare and Chris Winship for providing extracts of the March Current Population Survey from 1967 to 1988.

Income, Employment, and the Support of Children

At best parental income and work are only indirect indicators of children's well-being. Unlike children's health or education, neither parental income nor parental work are characteristics of children themselves. Nonetheless most people expect these characteristics of parents to affect children's well-being. But because the linkage is uncertain and poorly understood, parental income and work have less face validity as indicators of children's well-being than, for instance, infant mortality or high school graduation.

The usefulness of these indicators depends on establishing either theoretical or empirical links between income and work on the one hand and more direct measures of children's well-being on the other. Almost everyone believes that as parental income increases, children's opportunities also increase. But it is not so clear that raising parental income from say, $10,000 to $15,000 makes children better off if the median parent's income simultaneously rises from $15,000 to $30,000. The likely effect of changes in parental employment are even more ambiguous.

This paper is divided into two parts. The first is on indicators of economic well-being and the second looks at parental work. Each section begins with a discussion about how children's well-being is related to the indicator. Each section then discusses available indicators and why we need additional indicators.

INDICATORS OF ECONOMIC WELL-BEING

On average poor children fare worse than rich children on nearly every measure of well-being about which we collect systematic data. Poor children weigh less than rich children when they are born and they are more likely to die in their first year of life. When they enter school, poor children score lower on standardized tests, and that remains true when they graduate. Poor children are also absent from school more often and have more behavior problems than affluent children. Poor adolescents are more likely to drop out of high school, to have a baby, and to get in trouble with the law than adolescents from affluent families. Young adults who were poor as children average fewer years of schooling, work fewer hours, and earn less than children raised in affluent families. As a result children raised in poverty are more likely to be poor as adults and are more likely to need public assistance. It is no wonder that every attempt to assemble indicators of children's well-being includes some measure of family income.

Social scientists have at least two models of the way parental income affects children's life-chances, which I will call the investment model and the "good parent" models. Investment models hold that parents are rational individuals who invest both time and money in their children's human capital. They do this especially by investing in their children's education, but also by purchasing health, good neighbors, and other "inputs" that improve children's future well-being (Becker 1981). All else equal, children raised in affluent families are more likely to succeed than those raised in poor families because rich parents can invest more in their children than poor parents. This model implies that parents' absolute purchasing power is what matters, because the importance of income derives from what it buys. If this model is correct, our goal should be to replace income measures with direct measures of the market "inputs" that contribute to success, unless income measures are very good proxies for these inputs. Although we are not certain what these inputs are, most people believe that children must at least be well fed, adequately housed in a safe neighborhood, and get adequate medical care in order to take advantage of social and educational opportunities. These may not be sufficient conditions for children's success, but they appear to be necessary.

"Good parent" models hold that low parental income affects children by affecting parents' ability to be "good" parents. There are two versions of this model. The "parental stress" version holds that poverty is stressful and that stress diminishes parents' ability to provide "supportive, consistent, and involved parenting" (McLoyd 1990). This in turn has an adverse effect on the socioemotional development of children, limiting their educational and social opportunities.

The "role model" version of the good parents model holds that because of their position at the bottom of the social hierarchy, low-income parents develop values, norms, and behaviors that cause them to be "bad" role models for their children. Since children often copy the behavior and values of their parents, these "dysfunctional" parental values and behaviors are often transmitted to their children. Role model hypotheses sometimes also hold that behavior which appears to be dysfunctional from the point of view of the middle class is a rational response to poverty. This is likely to be true for families experiencing long-term poverty who have adapted to their economic conditions. For families experiencing short-term poverty, parental stress may have a greater affect on parental behavior. According to the role model hypothesis, increases in parental income do not improve children's life chances in the short run, but do improve them in the long run since parental income only changes children's culture over the long run.

The good parent model suggests that income affects parents' psychological well-being, which in turn affects their ability to be good parents, which then affects children's well-being. This model implies that parents' relative economic standing, as well as their absolute level of economic resources, may be important to children's well-being.

The next part of this paper discusses trends in parental income. After that I discuss the degree to which income is related to families' living standards. Because this relationship is weak, it is important to provide information on direct indicators of the material living conditions that we believe are important for children in addition to measures of parental income. I do not try to estimate the degree to which parental income is related to parental stress or culture, partly because we do not have good data on the latter and partly because it is a much more complicated question than can be addressed in this paper. I use several data sets throughout this paper. All of these are discussed in the Appendix.

Income-Based Measures of Children's Economic Well-Being

Because tastes differ, income is not a good proxy for the particular goods and service that a family consumes. Some families prefer more vacations and a smaller home; others prefer a large home and fewer vacations. But most people think that income is a good proxy for families' overall command over resources, and most believe that families will usually purchase "necessities" before "luxuries". Therefore, most people believe that as parental income declines, the chances that children's basic material needs will be met also declines. This would surely be true if income really did reflect command over resources. But a family's annual money income is not, in fact, a very good measure of its command over resources. This is true for many reasons:

> *Reporting errors*. Many families seriously under-report their income in surveys, and some over-report. Errors are especially common at the top and bottom of the income distribution.
>
> *Taxes, borrowing, and saving*. Taxes, borrowing, and saving all vary substantially among families with the same income. As a result, families with the same annual income can spend quite different amounts during the year on goods and services.
>
> *Noncash transfers*. Even when families spend the same amount, the value of what they consume can vary because of differences in their ability to get free (or subsidized) goods and services. These noncash transfers can come from the government, from employers, or from friends and relatives.
>
> *Noncash assets*. Families command over resources also depends on what they already own. The "service flows" from owner-occupied housing and from automobiles bought at some time in the past play an especially important role in driving a wedge between living standards and current income.
>
> *Local price differences*. Because of local variation in prices, especially for housing, families that spend the same amount get more in some communities than in others.

Consumer efficiency. The efficiency with which a family spends its money also influences the price it pays for goods or services of any given quality. A skilled shopper can buy a better car for $5,000 or a better melon for $2 than an unskilled shopper.

Most people acknowledge these problems with using income to measure command over resources, but still assume that the correlation between income and command over resources is high. As a result almost all analysts agree that the economic situation of American children has deteriorated over the past twenty years.

Poverty. To support the claim that the economic conditions of children have deteriorated, social scientists and policy makers often cite the poverty rate of children. Most economic indicators, including poverty, are quite sensitive to the years one chooses to compare, since economic indicators are influenced by the business cycle. For most of the tables in this paper I use all of the years for which data are available beginning in 1969. In the case of the Current Population Survey, I show data for a selection of years as a convenience. Eventually all years should be included. Business cycle peaks (assessed as peak years of GDP growth) occurred in about 1969, 1973, 1979, and 1989.

The first column in Table 1 shows that the official poverty rate for children has increased since 1969 when 14 percent of children lived in poor families. In 1991 the official child poverty rate was 21.8 percent.

The official poverty rate has been criticized on many grounds (Mayer and Jencks 1989, Mayer 1993, Ruggles 1990). Some of the criticisms raise doubts not just about the incidence of poverty, but about trends as well. I will discuss only three of these criticisms here. First, the trend in poverty depends on how the poverty threshold is adjusted for changes in prices. Second, the level and trend in poverty depend on what data set one uses to calculate the poverty rate. Third, the adjustments for family size implicit in the poverty line have no theoretical or empirical justification.

The first column in Table 1 shows official child poverty rates between 1969 and 1991. Official poverty statistics compare each family's income to a poverty threshold developed by Mollie Orshansky in 1964. If a family's income falls below the threshold it is classified as poor. These thresholds have been adjusted for changes in prices using the Consumer Price Index for urban consumers (CPI-U). The CPI-U, like the other indices that are used to adjust income for changes in prices, embodies hundreds of arbitrary decisions and compromises, some of which introduce systematic upward or downward bias. If these biases persist, their cumulative effect can be substantial. Economists agree that the CPI-U over-stated the annual rate of inflation during the 1970s because of the way it computed housing costs. This problem was especially severe during the late 1970s when the cost of buying a new house increased faster than most other prices. In 1983 the error was corrected. But earlier poverty statistics were not revised to reflect this correction. Official poverty statistics therefore reflect better price adjustments after 1983 than before.

The Census Bureau also publishes an alternative poverty series which adjusts the 1967 poverty thresholds using the CPI-U-X1, in which a change in the treatment of housing is applied beginning in 1967. This alternative is shown in column 2 of Table 1. This is clearly a better indicator of poverty than the official poverty rate. It shows that the child poverty rate increased 4 percentage points between 1969 and 1989 compared to the 5.6 percentage point increase in the official poverty rate.

These child poverty rates are based on family income. The Census Bureau defines a family as everyone living in a single housing unit who is related by blood, marriage, or adoption. Thus, if a woman lives in her home (and is what the Census Bureau refers to as the "reference" person) with a boyfriend and their child, the mother and child are counted as one family and the father is counted as a separate unrelated individual. If the mother's income is below the poverty threshold for a family of two, she and her child are classified as poor, regardless of how much money her boyfriend makes. This distinction does not seem reasonable, and as rates of co-habitation increase it may increasingly distort the true economic well-being of children. The most obvious alternative is to calculate poverty rates based on household income rather than family income. A households includes all the people who live in a single housing unit, regardless of their relationship to one another.

Column 3 in Table 1 uses the CPI-U-X1 to inflate the 1967 poverty thresholds and substitutes household income for family income. If one thought that the 1967 poverty thresholds were correct, the estimates in column 3 would probably be the best available. This measure yields an increase in the child poverty rate of only 1.4 percentage points. But there is no reason to think that Orshanky's poverty thresholds were "correct". One alternative is to begin with the 1992 official poverty thresholds, which are about 8 percent higher than the 1967 thresholds in real dollars because they were adjusted with the CPI. We can then adjust these thresholds downward using the CPI-U-X1 to get the same real values for earlier years. These are shown in column 4 of Table 1. Using this measure we get an increase in child poverty of only 1.1 percentage points between 1969 and 1989.

All of the child poverty rates so far use data from the March Current Population Survey. The CPS is a major government survey, but it is not the only such survey. To test the sensitivity of the poverty rate to different data sources, the last column shows the child poverty rate calculated in the same way as column 4 but using data from the decennial Census. It shows a decline of 1.2 percentage points. Thus, the evidence that the child poverty rate has increased substantially over the last 20 years is highly sensitive to reasonable changes in the way we measure poverty. All measures show what most people would probably agree is a high current child poverty rate. But this estimate ranges from 21.8 percent to 17.1 percent in 1989.

To adjust income for differences in household size requires an equivalence scale that shows how much money households of different sizes need to be equally well off. The equivalence adjustment in the official poverty thresholds reflect neither a sound theoretical rationale nor empirical findings. In fact, no one equivalence adjustment makes families equally well off in all respects. Scales that try to equalize adults' subjective well-being require small adjustments for household size (Vaughn 1984, Rainwater 1974), while scales that try to equalize households' material well-being or consumption require larger adjustments (Lazear and Michael 1980, Van der Gaag and Smolensky 1981, Mayer and Jencks 1989).

The size elasticity implied by the poverty thresholds is about .85 for families of three or more. This means that a 100 percent increase in family size requires an 85 percent increase in income to maintain the same level of economic well-being. These adjustments appear to be slightly low for measures of material hardship (Mayer and Jencks 1989). But for cognitive test scores, teenage childbearing, single motherhood, and dropping out of high school the size elasticity is greater than one (Mayer 1995). This means that family income must more than double to offset the effects of doubling family size. It follows that doubling family size is more detrimental to children's life chances than decreasing income by half.

It is hard to believe that elasticities this large are due to the reduction in economic resources that accompany additional family members, since there are economies of scale for most of the goods and services that families consume. When a family doubles in size it does not need to double the space it occupies, the number of televisions or cars it owns, or the amount of food it buys. This is why the adjustment in the poverty line is less than one. If we want the poverty line to be a proxy for material well-being, the size adjustments of the poverty thresholds may be about right (Mayer and Jencks 1989). But if we mean for the poverty line to be a proxy for broader aspects of children's life chances, these adjustments may be too low. This implies that a family's size and its income should not be concatenated into one measure, such as a poverty rate, unless we are sure what we want to measure.

In this section I have shown that both the level and trend in children's poverty rates is sensitive to how poverty thresholds are adjusted for changes in prices and the data set that one uses to estimate poverty. Both the level and trend are sensitive to the income unit as well. Most people would agree that column 3 in Table 1 provides better estimates than the official poverty measure, but there is little agreement about which among the other columns is superior. Therefore, the only alternative if one wishes to include a measure of poverty among indicators of economic well-being, is to provide a range of poverty estimates.

A poverty rate, however accurate, tells us only what happened to children at the bottom of the income distribution. It does not tell us what happened to the average child. Nor does it tell us what happened to affluent children. Those who worry about children's well-being usually worry less about what happens at the top of the income distribution than what happens at the bottom, but if children's well-being depends largely on their relative economic standing rather than absolute economic position, trends at the top of the income distribution may affect children at the bottom.

Income of the Median Child. Table 2 shows trends in real household income of children (adjusted with the CPI-U-X1) using both CPS and Census data. The mean of the third quintile is approximately the median income, so both Census and CPS data show that the income of the median child's household increased during the 1970s and hardly changed during the 1980s.

These estimates make no adjustment for differences in household size. This strategy assumes that from a child's viewpoint the benefits of additional siblings (or having two adults in the household rather than one) exactly equal the costs. This is unlikely. The average size of children's households declined from 4.25 to 3.39 over this period. Table 2 also shows estimates of the per capita income of children's households. This measure implicitly assumes that there are no economies of scale in larger households. These two alternative adjustments for size presumably bracket the "true" equivalence scale. Again the trend is the same in both the Census and CPS, namely that median per capita income increased a lot in the 1970s and less in the 1980s. In both data sets the increase in per capita median income was much greater than the increase in unadjusted median income in both the 1970s and 1980s. Much of the improvement in real per capita income is thus traceable to declining household size rather than rising money income.

Regardless of the equivalence scale or the data set, the trend in median household income is the same: the real household income of the median child grew during the 1970s and grew at a slower rate during the 1980s.

The estimates in Table 2 use the CPI-U-X1 to adjust for prices. Like trends in the poverty rate, trends in the median child's household income are sensitive to the way we adjust income for changes in prices. In Census data the CPI-U-X1 suggests that the median child experienced a 9.7 percent increase in real household income between 1969 and 1989. When we use the CPI the median income of children's households hardly changed between 1969 and 1989. Many economists prefer to measure price changes using the implicit price deflator for Personal Consumption Expenditures (PCE) in the National Income and Product Accounts. It implies that the real income of the median household with children rose 6.7 percent. The implicit price deflator is difficult to interpret, however, because it does not describe the price of a fixed market basket of goods. The fixed-weight PCE index for the market basket that consumers bought in 1987 rose more slowly than the implicit price deflator. When we use this index the median household with children experienced a 15.3 percent increase in its purchasing power between 1969 and 1989.

Most economists who study these matters also believe that standard price adjustments underestimate the value of qualitative improvements in the goods and services that consumers buy. If this bias meant that the true rate of inflation was one point less than the fixed-weight PCE index implies, the purchasing power of the median households with children would have risen by 42 percent between 1969 and 1989.

Inequality. Table 2 shows that in the 1970s and the 1980s income unadjusted for household size grew for children whose households were in the top half of the income distribution and fell for those in the poorest fifth of the income distribution. But CPS data show that per capita income fell during the 1970s for those in the poorest fifth of the income decile while Census data shows that per capita income grew among this group. Consequently the two data sets yield quite different conclusions about the decline in per capita income in the bottom of the income distribution between 1969 and 1989. The CPS shows a decline of 9 percent in the poorest decile, but Census data show a slight increase.

Because income grew at the top of the distribution, inequality grew regardless of the data set. However, it is unclear whether income growth at the top of the income distribution hurts children whose household income failed to grow. That depends on whether we think that relative or absolute economic well-being affects children.

Most people rely on published CPS data for trends in inequality in household income. Published data almost always use the CPI to adjust for prices. Relying exclusively on this measure could provide misleading information about the growth in economic inequality among children. But there is no agreement on a single alternative that is any better. There is no apparent reason to believe that either Census or CPS data are superior to the other. Researchers disagree about the "correct" adjustment for household size and the "correct" adjustment for prices. Consequently, no single measure of income should be used as an indicator of households' economic well-being. Several indicators using different price adjustments and household size adjustments (and data sets when possible) will provide a fuller picture of children's economic well-being.

Annual Income Versus "Permanent" Income. All the estimates so far rely on measures of household income measured in only one year. Annual income has two components. The first is a relatively stable or "permanent" component, which ensures that income in one year is fairly highly correlated with income in other years. The second is an unstable or "transitory" component that keeps the inter-annual correlation below 1.00. Most economists believe that the transitory component of income has little effect on a family's living standard because when income is low, parents will borrow against future income or draw down savings from past income in order to consume at the level of their permanent income. If they wanted an indicator of children's "true" economic well-being, most economists would probably measure families' permanent incomes. Studies show that using several years of parental income increases the intergenerational correlation of income (Solon 1993, Zimmerman 1993). Other studies show that using only one year of income can seriously underestimate the relationship of parental income to high school graduation (An, Haveman and Wolfe), and to children's cognitive test score, teenage childbearing, and educational attainment (Mayer, 1995). This implies that trends in "permanent income" would be a better indicator of children's well-being.

Unfortunately, we have no way of actually measuring "permanent" income. I used the 1989 wave o the Panel Study of Income Dynamics to calculate children's parental income in 1968-72, 1973-77, 1978-82, and 1983-87. Table 3 compares trends in the distribution of these five-year income averages to trends in the distribution of income measured in the year in the middle of the interval. Comparing Table 3 with Table 2, one can see that mean income is higher in the PSID than in either the CPS or the Census. That might reflect either selective sample attrition or better reporting in the PSID. As is well known, there is less inequality in the five-year averages than in a single year. This is particularly true for inequality between the bottom and the middle quintiles. Because incomes fluctuate, the five-year averages for the bottom quintile is 18 to 33 percent higher than the amount received by those who fall in the bottom quintile for a given year.

Inequality grew more for income measured in a single year (the decline in both the "20/50" ratio and in the "20/80" ratio was greater) than for income averaged over five years. But the difference in the trend is too small to be of much interest. For this time period, trends in annual income appear to parallel trends in five year income averages. Nonetheless, because it may change in the future, indicators of long-term income may be beneficial.

Most people believe that when poor children get poorer, their material standard of living declines. Similarly, because the poverty line is supposed to represent a constant level of purchasing power, the fact that the poverty rate for children has increased implies that children's material well-being has deteriorated. Many people also believe that when children's material well-being deteriorates their chances for a successful life do the same As I discuss next, the relationship between income and material well-being is not as strong as many assume. If this relationship is not strong, the relationship between income and social or psychological well-being may also be weak, although that issue is beyond the scope of this paper.

Measures of Consumption

When we examine the past twenty years, we find that most groups' real income has changed less than one percent per year. Meanwhile, tax rates fluctuated substantially, saving rates fell, noncash transfers to children grew, and home ownership declined, especially among low-income families. In addition, more women worked, so consumers bought more goods and services in the marketplace and produced fewer at home. The efficiency with which consumers spent their money may also have changed. Taken together, these changes could well be more important to economic well-being than a 10 or 20 percent change in real income.

How much families consume is a better measure of their current living standard than how much income they have. I define consumption as expenditures for goods and services that are actually consumed by a household, excluding income that is either saved, taxed, or unreported.[1] If we are interested in children's well-being, we should perhaps focus exclusively on consumption that benefits children, but in practice there is no way of doing this, so I focus on the overall level of consumption in children's households.

The measure of annual consumption used here includes total cash outlays for all items except the following: taxes[2]; purchases of stocks, bonds, and other investments; pension contributions; down-payments and mortgage payments for owner-occupied housing; purchases of motor vehicles; interest; gifts. Consumption also includes the following noncash items: the estimated rental value of owner-occupied housing[3]; the estimated depreciation of motor vehicles[4]; and the estimated value of food bought with Food Stamps.[5]

This measure of consumption is not ideal, but it should tell us more about the resources available to support children in any given year than the household's reported money income does. Since tax rates, saving rates, food stamps, and the relationship of mortgage payments to the rental value of owner-occupied housing have all changed since the early 1970s, especially for the poor, the relationship between consumption and income may well have changed as well. If underreporting of income has risen more than underreporting of consumption, that should also show up in these data.

Table 4 shows that in the CEX, as in the Census and CPS, income declined in the poorest children's households during the 1970s and 1980s. But Table 4 also shows that low income households consume goods and services worth far more than their reported income. This remains true even when we eliminate households that failed to answer one or more income questions.[6]

The ratio of consumption to income rose during the 1980s for all income groups, but it rose much more for low-income than high income children's households. The poorest ten percent of all households with children reported consumption averaging 185 percent of their income in 1972-73, 197 percent in 1980-81, and 236 percent in 1988-90. Even if we restrict our attention to low income households that answered every income question, consumption rose during the 1980s while income fell.

Inequality in consumption also grew less than income inequality. Table 4 shows that the ratio of mean income in the poorest decile to mean income in the third quintile declined in the CEX as in the Census and CPS. But when we turn to consumption the ratio of the poorest decile to the third quintile did not decline.

The high ratio of consumption to income is partly due to the fact that consumption includes the reported value of food stamps, the estimated rental value of owner-occupied housing, and depreciation for vehicles purchased in earlier years, whereas income does not. But even when we exclude these amounts, low-income households report cash expenditures far higher than their income.

If permanent income is more highly correlated with current consumption than with current income, the best way to get a realistic picture of trends in consumption among the long-term poor is to classify families by their current consumption rather than their current income. Table 5 shows that households with very low consumption lost ground both during the 1970s and during the 1980s. But the decline during the 1980s was considerably smaller than the decline during the 1970s. It was also smaller than the income decline in the decennial Census or in the CPS.

We can also use Table 5 to estimate the overall level of inequality in consumption. The Census shows that the poorest tenth of all households with children reported incomes averaging 12 percent of the median in 1989. In the CEX, the poorest tenth of all households with children consumed goods and service worth 32 percent of those that the median household consumed in 1989-90. Annual consumption is thus far more equally distributed than annual income. That is probably a reflection both of the fact that some households are only temporarily poor, allowing them to consume more than they take in during a given year, and the fact that a large fraction of households with very low reported income also have unreported income. Inequality in consumption also grew more slowly than inequality in measured income. The ratio of mean consumption in the poorest decile to the mean in the third quintile fell from 37.9 percent in 1972-73, to 32 percent in 1980-81 and remained at that level in 1989-90.

Comparing consumption to income shows that:

> Consumption of the median child's household increased between 1972-73 and 1980-81, but declined slightly between 1980-81 and 1989-90. Median income also increased during the 1970s, but it stayed about the same during the 1980s.

> Among low-income households consumption increased during the 1980s but income declined.

> Consumption is much more equally distributed than income.

> The lowest consuming children's households consumed less in 1989-90 than in 1980-81, or 1972-73, but the decline in consumption was less than the decline in income over the same period.

> Inequality of consumption increased modestly during the 1970s and not at all during the 1980s, while income inequality increased in both decades.

Because trends in consumption do not parallel trends in income and because consumption is probably a better measure than annual income of command over resources, indicators of children's well-being would ideally include measures of their household's consumption. Unfortunately, CEX data are the only source for data on consumption and they are difficult to use. In addition BLS's treatment of missing data and its sample selection procedure raises important questions about the usefulness of these data for this purpose. (See footnotes and Appendices.)

Material Well-Being

Once we recognize the wide range of uncertainty about the true rate of inflation and allow for the possibility that taxes, saving, borrowing, noncash benefits, noncash assets, consumer efficiency, and need may have changed substantially during these years, it becomes easy to imagine that children's material well-being might not track trends in household income very closely.

Since consumption declined less than income among those with low income, material well-being is unlikely to have declined by as much as income statistics suggest. Households build up stocks of goods that reduce their need to spend money. Thus the service flows from a purchased home, furniture and other durables improve material well-being without additional expenditures. Furthermore, government effort on behalf of poor children has been aimed at reducing their material deprivations through noncash transfers.

Ideally one would like a single measure of living conditions analogous to measures of income that allowed us to say that one child lives twice as well as another. To do this requires measuring all the important living conditions and weighting them by their relative importance. Unfortunately we do not have data on all of the measures of living conditions that children might consider important. Judging by government expenditures, most citizens believe that adequate housing, food, and medical care are more important than anything else. By using a combination of data sets, it is possible to get trends on housing conditions and access to medical care. It is also possible to get trends on whether children's families own some common consumer durables. But no national data set includes good information on food consumption. Absent measures of all of the important living conditions, one could collect information on a random sample of goods and services. But no data set does this either. Furthermore, no set of weights exists for creating a single measure of living conditions.

In this section I assess whether trends in children's material well-being parallel trends their parent's income. This section serves two purposes. First, if these trends do not parallel one another, then trends in income cannot be used, as many believe that they can be, to infer trends in children's material well-being. Second, many of the material "hardships" are themselves important indicators of children's well-being with considerable face-validity: we should know how many children live with serious housing inadequacies or other material deprivations.

Housing. The last column in Table 6 shows trends in the percent of children living in homes with various problems for which data are available. The first part of the table focuses on problems with the dwelling unit itself. The percent of children living in homes without a bathroom, with no sewer or septic tank, with no central heat, and with too few electrical outlets declined by at least a percentage point between 1973-75 and 1985-89. Maintenance problems (hole in the floor, cracks in the walls or ceiling, leaky roofs) increased, but the increase was always less than 1 percent. The percent of children living in crowded households declined. Even the percent of parents who reported that crime is a problem in their neighborhood was slightly lower in 1985 (the last year for which data were available) than it had been when the survey started in 1973.

The fact that most of these housing conditions improved a little is not surprising since the household income of the average child increased a bit over this period. As one would expect, low-income children are also more likely than the average child to experience all these housing problems. This might reinforce the notion that indicators of children's household income are sufficient to infer trends in children's material well-being. But even among children whose household income is low (the bottom decile) whose real household income declined, almost all of these housing problems became less common. The exception to this rule is cracks in the wall or ceiling. But this problem increased for middle class children as well. Low-income children were also more likely to live in rented housing, but so were middle class children.

Table 7 shows the percentage point difference between the bottom decile and the middle quintile of children's households on each of our measures in different years.[7] The last two columns show whether this difference widened (+), narrowed (-), or stayed the same (o) in the 1970s and 1980s.

The gap between the middle and the bottom narrowed for almost all maintenance problems and design inadequacies. But neighborhood crime increased for low-income children relative to middle class children and the gap between children at the bottom and children in the middle also widened for home ownership. Taking the period from 1970 through 1990 as a whole, therefore, material inequality between the bottom and the middle seems to have declined somewhat, even though income inequality between the bottom and the middle was increasing.

Medical Care. Table 8 extends this analysis to a different domain, medical care. It shows the percent of children with no doctor visit in the previous year and the number of doctor visits for children with at least one visit. It shows these estimates separately for children under seven years old and those seven to seventeen years old, since the medical needs of the two age groups may be different. The HIS was changed in 1982 in ways that affect these estimates. Therefore the estimates for 1970 and 1980 are comparable to one another, but not to the estimates for 1982 and 1989.

Because the HIS only asks about the parents' broad income category, not their exact income, I could not identify the bottom decile at all precisely. I therefore estimated these children's doctor visits indirectly. First, I regressed annual doctor visits on the natural logarithm of family income, holding family size constant. Then I used the Census data in Table 2 to estimate the income differential between the average parent and parents in the bottom income decile. Finally, I combined these two estimates to predict the frequency with which children from very low income households visited the doctor.

Table 8 shows that like most other resources available to children, their doctor visits increased during both the 1970s and 1980s. The likelihood of visiting a doctor increased among low income children even more than for children in general between 1970 and 1980.[8] But from 1982 to 1989 the increase is slightly smaller for low-income children than for more affluent children.[9] The trends are similar for the number of doctor visits. Low income children's access to physicians did not deteriorate, as one might have expected given the reduction in both their parents' overall purchasing power and in insurance coverage.[10]

Consumer Durables and Telephone Service. Table 9 shows several additional measures of material well-being. Some of these, such as dish washers and air conditioning, might be considered "luxuries". Others like having a telephone might be considered necessities. Because parents' tastes vary, some parents will choose to forgo air conditioning in favor of a dishwasher and others will have the opposite preference. But if parents purchase goods and services in the order of their importance, families that have dishwashers are also likely to have other more basic material resources.

Table 9 shows that children's households became more likely to have all of these items except clothes washers between the early 1970s and the late 1980s. Given the trend in clothes dryers and dishwashers, we suspect that the data for clothes washers is inaccurate, but we have been unable to discover any reason for this. Poor children's households also became less likely to have at least one motor vehicle. However, their likelihood of having two or more vehicles increased. This implies that the bottom income decile includes more very poor households, but that it also includes more "mistakes". The improvement for low income children was greater than for children in general. On these "luxury" items, poor children apparently became more like middle class children.[11]

Food. We have no good national time series data on children's food consumption or nutrition. USDA has conducted the National Food Consumption Survey at about ten year intervals since 1955. But the last survey in the mid-1980s had such a low response rate that the data are unusable. The CEX includes information on what households spend on food. Many economists believe that as economic well-being increases the proportion of income that households will allocate to necessities will decline, while the proportion that they allocate to luxuries will increase. CEX data show only a small and statistically unreliable change in the proportion of consumption that goes for food in children's households or even in low-income children's households between 1972-73 and 1988-90.

Another approach to assessing the adequacy of food consumption is to ask parents how often their family goes without the food it needs. We have no time series data on this question, but the Survey of Income and Program Participation asked it in a topical module. These data show that 4 percent of children's parents reported that their family sometimes or often did not have enough to eat. In the poorest income decile, the number is 12.4 percent. It is almost as high, 11.6 percent, in the second poorest income decile. Almost all these parents say that the reason they did not have enough to eat was because they did not have enough money.

Conclusions about Indicators of Children's Economic Well-Being

No single measure can provide reliable evidence about changes in children's overall economic well-being. As long as we are interested in what has happened to the economic well-being of the average child, many income measures produce a relatively consistent story, and that story is in turn consistent with trends in both consumption and the measures of material well-being for which we collect data. All measures seem to suggest that the economic well-being of the median child improved over the last two decades. However, the degree of improvement is sensitive to 1) the method used to adjust for prices, 2) the adjustment for household size, 3) the data set used for the estimates, and 4) whether we use income, consumption or material well-being as an indicator. To the extent that the degree of change, rather than the direction of change, is important multiple indicators will be needed to produce reliable information.

If we are interested in the distribution of economic well-being, different measures and data sets can produce quite different conclusions. Conclusions about the growth in income inequality depends on the adjustments for household size. Trends on income inequality do not mirror trends in inequality of consumption. Trends in the measures of material well-being for which we have consistent measures also do not reflect trends in income inequality. This strengthens the argument for multiple indicators of economic well-being.

If we want to know about trends in children's housing, health care, food consumption, or other material conditions, we must measure these trends directly, rather than assuming that changes in children's money income predicts changes in material well-being. Some measures of material hardships are available in nationally representative data sets, but they are seldom published in a way that makes them useful. Furthermore, these measures of material well-being are collected in different surveys, so we cannot currently tell how many children live with multiple material deprivations. Data on "food hardships" are especially scarce, even though data from SIPP suggest that this is not a rare problem.

How important either income or material living conditions are to children's well-being is an open question.

INDICATORS OF PARENTAL EMPLOYMENT

Although almost everyone agrees that as parents' income increases so do their children's life chances, there is no such consensus about how to interpret trends in parents' work. Americans have always believed that fathers should work. But they are much more ambivalent about whether mothers should work. When parents work two things happen: their income increases, but the time that they have available to devote to their children and to home production decreases. Some people focus on the "time effect". They are alarmed at the increase in mothers' labor force participation because they fear that children whose mothers work are less likely than children whose mothers stay at home to be adequately supervised and nurtured. Others focus on the "income effect". They are alarmed at mothers who do not work enough to earn the money it takes to buy the things their children need.

Children will presumably benefit from their parents' working if the benefits of increased income out weigh the loss of their parents' time. Consequently, almost everyone believes that at least one parent should work in a two parent family, because the added income will out weigh the benefit of having two rather than one parent at home. Whether the second parent should work is a more complicated question, since it is not clear that the added income will out weigh the costs to children of having no parent at home for much of the day. The same question arises with respect to single parents. Once children are school age, they are without their parents for several hours a day whether parents work or not, so most people think that as children get older any harmful effects of working parents will diminish. If the family is poor without the second parent working, the income effect might out weigh the time effect. These arguments imply we should provide indicators of parental work separately by parents' marital status and the age of children.

The benefit to children of parental work depends on the relative quality of care provided by parents and those who care for their children in their absence. If the care a child gets when her mother works is worse than the care her parents could have provided, this might hurt the child. If the care is better, it might help. If the parental characteristics that benefit children are the same ones that employers value, the care that children get in their parents' absence will be inferior to the care they would have received from their parents, because parents will not pay more for childcare than their own wage. Non-parental childcare might be as good as parents' care if the characteristics that are of value to children are not the same as those valued by employers. Non-parental care might also be better if either the childcare or wages of low-skill workers is subsidized, because then they could in principle purchase childcare at a price higher than their own wage. Neither theory nor empirical research on the relative merits of non-parental and parental childcare provide strong evidence about how to interpret trends in parental work. But they do suggest that in order to evaluate these trends, indicators of parental work should be accompanied by indicators of what their children do in their parents' absence, but this is beyond the scope of this paper.

The focus on the income gains from market work may be misleading for two reasons. First, to get an adequate picture of the economic benefits of market work, the economic gains from market work must be discounted by the value of loss home production (as well as the monetary costs associated with work). Many studies have shown that home production has an important impact on economic well-being (Gottschalk and Mayer 1994). Gronau (1980) estimated that in 1973 among white married-couple households the value of home production was equal to 70 percent of households' money income after taxes. His estimates show that among households with young children, the loss of home production when the wife joins the labor force almost equalled her increased money earnings.

Second, additional income may not benefit children as much as it benefits adults. Some evidence suggests that a greater proportion of each additional dollar that a family gets will be spent on a child when family income is low than when it is high (Lazear and Michael 1988). As incomes increase, additional income is "frosting on the cake", with greater benefits to adults rather than children. Thus a greater proportion of the earnings of middle class mothers will go to "luxuries" that provide no direct benefit to children. Since single mothers are more likely to be poor if they do not work, some people seem to think that the benefits of additional income will out weigh the loss in time that single mothers spend with their children.

Some social scientists (e.g. Meade 1986) and policy makers have suggested a potentially important non-monetary benefit of work, namely that work builds character. For instance, those who advocate work requirements for welfare recipients contend that work not only increases income, but that it also reduces the depression, alienation, and lethargy that result from welfare dependency. According to this reasoning, working parents are better for children than welfare-dependent parents because they provide better role models and because they are happier and more socially integrated. Some people also believe that this is also true for middle class mothers. The evidence that work is improving for either middle class or poor parents is sparse, however, so it provides little guidance about how to interpret trends in parents' labor market work.

Research on the effect of parental work on children's socioemotional development, cognitive skills, and educational achievement (Chase-Lansdale 1994, Heyns 1982) is contradictory. A few studies, suggest that the income might have a greater effect on low-income than high-income children's educational achievement (Hoffman 1980, Milne et al. 1986) and cognitive skills (Desai et al. 1989). But other studies produce contrary results (e.g. Heyns 1982, Heyns and Catsambis 1986). Empirical research is still too inconclusive to provide a strong rationale for thinking that increases in parents' labor market work is either a benefit or liability to children. Nonetheless, the increase in maternal work has been one of the most important changes in the family over the last twenty years. Therefore, providing consistent indicators of parental work may be useful. These indicators should capture the two implicit concerns of those who worry about parents' work, namely the time that parents have available for their children and the income that they earn from work. I discussed income (though not wages) in the last section, so this section will concentrate on the first of these concerns.

Indicators of Parental Work

Labor Force Participation. The most common indicator of parents' employment is the labor force participation rate. But it may not be the best indicator. The labor force participation rate counts everyone who is working or unemployed as a "participant" in the labor force. An individual is counted as unemployed if he or she was not employed but made specific efforts to find employment within the previous four weeks. Therefore, the participation rate includes both people who are working and people who are not. It also gives equal weight to each adult. But mothers with many children are less likely to work than mothers with only one child. The participation rate tells us how many mothers and fathers participate in the labor force, but not how many children have mothers or fathers who are participants. Table 10 uses CPS data to shows trends in the percent of children whose father's and mothers were in the labor force by the marital status of their parents.

Like trends in income, trends in labor market work are sensitive to the business cycle. This table shows changes in employment status between 1970 and 1980 and between 1980 and 1990. These years are at fairly comparable points in the business cycle.

As is well known, labor force participation has increased a lot among mothers and changed little among fathers, and the increase among mothers has been greater among those who are married than among those who are single. The increase in labor force participation among mothers leads many people to believe that children receive less supervision and attention at home than they used to. The increase in labor force participation among married mothers leads many people to believe that poor single mothers should have increased their work effort more than they have.

Employment. Table 10 shows the unemployment rate by parents' marital status. The unemployment rate for married fathers was 2.6 in 1970, but it was 4.7 in 1980. Although 1990 was also a year after a peak, it was a weak recession and married fathers' unemployment rate was 3.8 percent. Thus married fathers' unemployment rates appear to have increased slightly. Unemployment rates for single mothers also increased.

Table 10 also shows employment rates, that is the percent of each group that is employed. This is a better indicator than the participation rate of how many parents actually work. Employment rates have declined more than labor force participation rates for fathers. Although the participation rates of single mothers has increased somewhat, the proportion of single mothers actually working has not increased.

Hours of Work. Among parents who work, some work part-time and some work full-time, so employment status is not a good indicator of trends in either economic support or time with children. The first part of Table 11 shows the number of hours worked by all parents in a week whether they worked or not. Because it concatenates changes in employment with changes in the hours that the employed work, it can be roughly interpreted as the trend in the hours that parents were unavailable for childcare and home production. The increase in total hours actually worked by mothers is more modest than one might imagine from changes in the labor force participation rate.

The second part of Table 12 shows the number of hours worked in a week among parents who worked. This is an indicator of changes in work patterns among workers. It shows that among workers the number of hours worked has hardly changed. Thus the increase in work hours among married mothers is mainly attributable to an increase in employment.

Because some people believe that parental work is more harmful for very young children than for older children, Table 12 shows the employment status and hours worked for parents of children 0 to 3 years old. As is well known, hours of work among married mothers of very young children has increased at a faster rate than hours of work among mothers of older children. The trend in hours worked among workers is similar for all married mothers and married mothers of very young children.

Time Available to Children. The total parental hours available to children is a function of how many parents are in the family and how much they work. The proportion of households with children with only one adult has increased. These households have fewer total hours to devote to both market work and home production, including child care, than households with more than one adult. In additions, mothers in married-couple families have increased their hours of labor market work, so they have presumably decreased the hours that they spend in home production and child care.

Although we do not have good data on the number of hours that parents actually spend caring for their children or otherwise producing goods and services in the home that benefit their child, we can estimate changes in the time available to children by subtracting the amount of time that parents spend at work from the number of non-sleeping hours in the day (sixteen per parent). Thus a child living in a married-couple family in which one parent works eight hours a day five days a week and the other parent does not work outside the home has in principle 7[16*2]-80=144 hours to spend with their children and in other forms of home production that benefit the child. A single parent who works full-time has in principle 112-80=32 hours per week to spend with her children and in other forms of home production.

These trends in non-working hours are only an approximation of the hours parents really do spend with children. They do not take into account time spent commuting, in leisure, or in other activities that may not benefit their children. They also do not take into account the quality of care that children get in their parent's absence. Table 13 shows that the amount of time that parents could in principle spend with their children or in home production that benefits their children declined by 17.2 hours from 1970 to 1992. But because the number of children in families also declined, the time available per child increased by 7.9 hours per week. Whether trends in overall time or time per child are more relevant to children's economic well-being is an empirical question. But the fact that parents actually had more time available per child in 1992 than in 1970 is contrary to what some people have inferred from labor force participation rates alone.

The decrease in total time available to children under three years old was similar to the change for children in general. But there was almost no change in the per child hours available to children under three years old.

If parents reduced their leisure to make up for the time that they now spend in the labor market (or to make up for the absence of a spouse), these numbers may over-state children's loss of parental time. Data on home production from the PSID in Table 14 show that parents spent 159 more hours per year in the market in

1987 than in 1976. But the hours that they spent in home production per year declined by 264 hours over this same period. This means that "leisure" increased by 105 hours over this period.

This data is consistent with evidence from time use surveys in the US reviewed in Juster and Stafford (1991). Between 1965 and 1981 women reduced their hours of home production by an average of 11.3 hours per week (from 41.8 to 30.5). However, the number of hours spent in the labor market increased by an average of only 5.0 hours (from 18.9 to 23.9 hours per week). The result was a net decline of 6.3 hours per week devoted to combined home and market production (or a 6.3 hour increase in leisure). Time-use studies also show that while the number of hours men devote to home production is roughly a third of the number of hours women devote, men did increase their hours of home production by an average of 2.3 hours per week (from 11.5 hours to 13.8 hours) between 1965 and 1981. On average they decreased market work by 7.6 hours per week, resulting in a net increase of 5.3 hours per week devoted to leisure.

Because of the large increase in women's labor market participation in the last twenty years, most people think that the time that parents have available for their children has declined, especially since more mothers remain single than twenty years ago. But the time that parents have to spend with each child has not declined. In fact parents have increased the hours that they spend in "leisure". If we are interested in parental work as a proxy for parental time available to children, labor force participation rates may be misleading. Indicators of parent's work must be balanced with indicators of what parents do when they are not working and the number of children that they have to care for. They should also be accompanied by indicators of what children do when they are not with their parents.

APPENDIX

Following is a description of the data sets used in this paper. In all data sets the unit of analysis is a child. All data sets are purportedly representative of the non-institutionalized population of the United States when appropriate weights are provided. In all cases these sampling weights were used.

The Decennial Census. The 1970, 1980, and 1990 Censuses collected information on whether each unit was occupied by the owner, how many rooms it had, how many people lived there, whether the unit had complete plumbing, how old the building was, whether the residents had a car or truck, and whether they had telephone service. The Census definition of "complete" plumbing became slightly more restrictive over time, so our estimates understate the true improvement.

The Current Population Survey. Many discussions of change in the distribution of income rely on data collected by the Census Bureau's Current Population Survey (CPS). The CPS currently surveys about 60,000 households a month and succeeds in interviewing someone in 96 percent of them. Every March the CPS asks detailed questions about household members' income from earnings, assets, and transfer payments during the previous calendar year.

The American Housing Survey (AHS). The AHS was conducted annually from 1973 through 1981 and biennially starting in 1983. In order to keep our work manageable, we work only with data it collected in odd-numbered years, plus that collected in 1974. It collects data on whether housing units have central heating, air conditioning, a modern sewage system (either a septic tank or a sewer hookup), whether there were electric outlets in every room.

The Consumer Expenditure Survey (CEX). The CEX surveyed national samples of consumers in 1972-73, 1980-81, and 1984-90. During the 1980s about 400 households entered the CEX every month. These households were surveyed four times and then rotated out. To get stable estimates we pool data for several adjacent years wherever possible. Except in 1980-81, the CEX asked consumers whether they owned a clothes washer, a clothes dryer, and a dishwasher.

The CEX tries to collect expenditure data from a representative sample of housing units four consecutive quarters. But it does not follow households when they move, so it gets a full year of data on a given family only if that family remains at the same address for the full year. The analyses in this paper omit households with less than four quarters of data. That means they are not fully representative of the population from which they are drawn, and that we cannot draw conclusions about population trends from the CEX data. We can, however, draw tentative conclusions about changes in the relationship between income and consumption, at least for those who do not move.

The CEX income data are also unusual. The Bureau of Labor Statistics (BLS) sets missing values to zero, rather than imputing a value from other respondents with similar characteristics. We can exclude these households in the 1980s, but in 1972-73 we can only identify such households if they failed to report any major source of income. We therefore present two different time series. The first includes everyone who reported a major source of income and sets missing values to zero in all years. This series covers both 1972-73 and the 1980s. Our second series is restricted to households that answered all the income questions. This series begins with households leaving the survey in 1981 and runs through 1990. These comparisons cover what BLS calls "complete income reporters" -- a term that has always included any consumer unit that reported any major source of income.

The Health Interview Survey (HIS). The HIS collects data on how often parents had taken each child to the doctor during the previous year, how many days they had kept each child in bed during the past two weeks, and how many days the parents had restricted each child's usual daily activities due to illness. Because the HIS interview schedule was revised in 1982, post-1982 HIS data are not comparable to pre-1982 data. I therefore report trends from 1970 to 1980 and from 1982 to 1989.

The Panel Study of Income Dynamics. The PSID is an ongoing longitudinal survey of US households begun in 1968 by the Survey Research Center of the University of Chicago. Low-income families were over sampled, but weights are developed to compensate for over-sampling and sample attrition. I use sample weights in all analyses. I use the 1989 wave of the PSID.

Notes

[1] The CEX actually provides data on what it calls "consumer units," not households. Consumer units are groups of individuals who live in the same household and either (a) are related to one another or (b) pool resources to purchase any two of the three categories of goods and services about which BLS inquires (food, housing, and "other expenses"). But since only two percent of households contain more than one consumer unit, we use the terms interchangeably in the text.

[2] Our estimate of federal income tax liabilities differs from that on the public use data tapes in the 1980s. According to John Sabelhaus, who computed the values we used, the public use data tapes systematically understate federal income tax liabilities in the 1980s. Sabelhaus recomputed tax liabilities using the CEX income data. Because BLS includes sales taxes in the purchase price of specific items and does not record the state in which a CU lived on public use data tapes, we were unable to exclude sales taxes.

[3] Our estimate of a home's rental value is also taken from Sabelhaus's data tape. For owner-occupied housing this estimate is based on the owner's estimate of the home's market value, which was multiplied by the ratio of aggregate rental value to aggregate market value for all owner-occupied housing in the relevant year. According to Sabelhaus, the numerator of this ratio came from the National Income Accounts, while the denominator came from the Flow of Funds accounts.

[4] Our estimate of vehicle depreciation is based on a regression equation that estimates a consumer unit's mean annual expenditure (in 1989 dollars) for purchases of motor vehicles. The independent variables in this equation were the household's total expenditures on other forms of consumption (which predicts vehicle expenditures better than income does) and the number of vehicles that the consumer unit owned. We used these predicted values to smooth out year-to-year fluctuations in vehicle owners' outlays. In principle, we should have done the same thing with other durables, such as furniture, refrigerators, and stereo equipment, but the required data were not available in most years.

[5] Because of data limitations, consumption does not include the value of other noncash transfers, such as federal housing subsidies, employer-financed health insurance, Medicare, Medicaid, or free childcare.

[6] Eliminating households with incomplete income data has relatively little impact on the results for the bottom two income deciles because most of these households answer all the income questions.

[7] An alternative strategy for comparing the relative well-being of different kinds of households is to divide the decile mean by the grand mean rather than subtracting one from the other. But when the outcome is dichotomous, this approach often yields different answers when one measures the presence of a resource than when one measures its absence. Because of this problem, most analysts prefer to analyze differences in dichotomous outcomes by dividing the odds of one group's having a given outcome by the odds of the other group's having it. Like the arithmetic difference between proportions, odds ratios yield the same results regardless of whether one counts people with or without an attribute. But when the base rate is very high or very low, a small absolute difference between two groups can translate into a very large difference in odds ratios. Changes in odds ratios are therefore unlikely to have a linear relationship to any plausible utility function. That problem is not always solved by using arithmetic differences, but it is usually lessened. For a fuller discussion of this issue see Mayer and Jencks (1993).

[8] The decline in the coefficient of family income is statistically significant for both children under seven and children between the ages of seven and seventeen.

[9] The apparent increase in income's effect on the number of doctor visits is not statistically significant, but it recurs for both age groups, so it may be real.

[10] In previous work, we have found that the number of days that children limit their activity due to illness has increased (Mayer and Jencks 1993). But even when health status is taken into account the association between income and doctor visits increased for children between 1982 and 1989. For a more detailed analysis

of doctor visits in 1980 that controls self-reported health status, the presence of acute and chronic conditions, and bed days in the past year, see Mayer (1993).

[11] Although not shown in this paper, the conclusions about material well-being are qualitatively the same regardless of the adjustment for household size.

REFERENCES

An, Chong-Bum, Robert Haveman and Barbara Wolfe. 1992. "The Windows Problem in Studies of Children's Attainments: A Methodological Exploration." mimeo.

Becker, Gary. 1981. A Treatise on the Family. Cambridge: Harvard U. Press.

Chase-Lansdale, P. Lindsay. 1994. "Families and Maternal Employment During Infancy" in Exploring Family Relationships with Other Social Contexts edited by Ross Parke and Sheppard Kellam. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coder, John. 1991. "Exploring Nonsampling Errors in the Wage and Salary Income Data from the March Current Population Survey," Washington: Housing and Household Economic Statistics Division, US Bureau of the Census, offset.

Council of Economic Advisors. 1994. Economic Report of the President Government Printing Office.

Desai, S. and P. Lindsay Chase-Lansdale, and Robert Michael. 1989. "Mother or Market: Effects of Maternal Employment on the Intellectual Ability of 4-Year Old Children." Demography, 26:545-561.

Duncan, Greg, Timothy Smeeding, and Willard Rodgers. 1992. "Why Is the Middle Class Shrinking?" Ann Arbor: Survey Research Center, University of Michigan.

Goedhart, Theo, Victor Halberstadt, Arie Kapteyn, and Bernard van Praag. 1977. "The Poverty Line: Concept and Measurement." Journal of Human Resources, 12:502-520.

Gottschalk, Peter and Susan E. Mayer. 1994. "Changes in Home Production and Trends in Economic Inequality", mimeo.

Gronau, Ruben. 1980. "Home Production: A Forgotten Industry". Review of Economics and Statistics: 408-415.

Heyns, Barbara. 1982. "The Influence of Parents' Work on Children's School Achievement," in Families that Work: Children in a Changing World edited by Shelia Kammerman and Cheryl Hayes, Washington, D.C.: National Academy Press.

Heyns, Barbara and Sophia Catsambis. 1986. "Mother's Employment and Children's Achievement: A Critique". Sociology of Education, 59(July):140-151.

Hoffman, Lois. 1980. "Effects of Maternal Employment on the Academic Attitudes and Performance of School-aged Children". School Psychology Review, 9:319-35.

Juster, F. Thomas and Frank Stafford. 1991. "The Allocation of Time: Empirical Findings, Behavioral Models and the Problems of Measurement". Journal of Economic Literature, 24(2):471-523.

Lazear, Edward and Robert Michael. 1980. "Family Size and the Distribution of Real Per Capita Income." American Economic Review, 70:91-107.

Lazear, Edward and Robert Michael. 1988. Allocation of Income within the Household. Chicago: University of Chicago Press.

Mayer, Susan E. 1991. "Are There Economic Barriers to Seeing the Doctor?" University of Chicago, Harris School of Public Policy Studies, offset.

Mayer, Susan E. 1993. "Living Conditions among the Poor in Four Rich Countries." Journal of Population Economics 6: 261-86.

Mayer, Susan E. and Christopher Jencks. 1989. "Poverty and the Distribution of Material Hardship." Journal of Human Resources, 24: 88-114.

Mayer, Susan E. and Christopher Jencks. 1993. "Recent Trends in Economic Inequality in the United States: Income versus Expenditure versus Material Well-Being," in Dimitri Papadimitriou and Edward Wolfe, editors, Poverty and Prosperity in America at the Close of the Twentieth Century London: Macmillan, pp 121-203.

McLoyd, Vonnie. 1990. "The Impact of Economic Hardship on Black Families and Children: Psychological Distress, Parenting and Socioemotional Development". Child Development, 61: 311-346.

Meade, Lawrence. 1986. Beyond Entitlement: The Social Origins of Citizenship. New York: Free Press.

Milne, Ann and David E. Meyers, Alvin Rosenthal, and Alan Ginsburg. 1986. Sociology of Education, 59(July):140-151.

Rainwater, Lee. 1974. What Money Buys: Inequality and the Social Meaning of Income. New York: Basic Books.

Ruggles, Patricia. 1990. Drawing the Line. Washington, D.C.: Urban Institute Press.

Solon, Gary. 1992. "Intergenerational Income Mobility in the United States". American Economic Review June:393-408.

U.S. Bureau of the Census. 1993. "Measuring the Effect of Benefits and Taxes on Income and Poverty: 1992," Government Printing Office, 1993.

van der Gaag, Jacques, and Eugene Smolensky. 1981. "True Household Equivalence Scales and Characteristics of the Poor in the United States." Review of Income and Wealth, 28:17-28.

Vaughn, Denton. 1984. "Using Subjective Assessments of Income to Estimate Family Equivalence Scales: A Report on Work in Progress." Office of Research, Statistics, and International Policy, Social Security Administration, Washington, D. C.

Vickery, Claire. 1977. "The Time-Poor: A New Look at Poverty". Journal of Human Resources, 12(1):26-48.

Zimmerman, David. 1992. "Regression Toward Mediocrity in Economic Stature". American Economic Review, June:393-408.

Table 1

Trends in the Poverty Rate for Children Under 18 Years Old

	Published CPS Estimates		Estimates from Public Use Tape		
	Family Income		Household Income		
			1967	1992	
INCOME			Threshold	Threshold	
	CPI	CPI-U-X1	CPI	CPS	Census
1969	14.0	13.8	15.1	17.1	18.3
1979	16.4	16.4	14.9	16.6	16.2
1989	19.6	19.6	16.5	18.2	17.1
1991	21.8	20.0	28.3	20.2	--
Change:					
1969-89	5.6	4.0	1.4	1.1	-1.2

Source: Published CPS estimates are from US Bureau of the Census, "Poverty in the United States: 1992" Current Population Reports. Series P-60-185. Washington, D.C.: Government Printing Office, Table 3. Estimates from public use tapes were tabulated by David Knutson using data described in the Appendix.

c:\kids\irpkid.\4/30/95 (HOUSEHOLDS WEIGHTED BY KIDS)
SOURCE: KNUTSON, Cenpov.fam and cenpov.hhd 11-7-94

Table 2

Mean Income in 1992 Dollars for Children's Households, by Income Decile or Quintile and Year

	Decile		Quintile			
Year	First	Second	Second	Third	Fourth	Fifth
Household Income						
Census						
1969	5,937	15,801	25,154	34,878	45,610	80,998
1979	5,241	14,714	25,293	37,883	51,095	88,724
1989	4,722	13,737	24,754	38,265	54,104	101,460
Percent Change:						
1969-79	-11.7	-6.9	.1	8.6	12.0	9.5
1979-89	-9.9	-6.6	-2.1	1.0	5.9	14.4
1969-89	-20.5	-13.1	-1.6	9.7	18.6	25.3
CPS						
1969	7,985	16,976	25,425	34,715	45,250	73,945
1979	6,165	14,835	25,015	37,377	50,438	81,349
1989	5,504	13,157	23,472	37,207	53,257	91,611
Percent Change						
1969-79	-22.8	-12.6	-1.6	7.7	11.5	8.7
1979-89	-10.7	- 9.9	-6.2	.5	5.6	12.6
1969-89	-31.1	-22.5	-7.8	8.2	17.1	22.3
Per Capita Income						
Census						
1969	1,075	2,800	4,643	6,810	9,378	13,210
1979	1,147	3,157	5,490	8,304	11,526	17,312
1989	1,092	3,033	5,489	8,683	12,592	20,695
Percent Change:						
1969-79	6.7	12.8	18.2	21.9	22.9	31.0
1979-89	-4.8	- .4	.0	8.2	9.1	19.5
1969-89	1.6	8.3	18.2	27.5	34.3	56.6
CPS						
1969	1,418	2,992	4,687	6,752	9,242	15,666
1979	1,351	3,159	5,489	8,298	11,506	18,833
1989	1,290	2,923	5,314	8,623	12,562	22,399
Percent Change						
1969-79	-4.7	5.6	17.1	22.9	24.5	20.2
1979-89	-4.5	-7.4	-3.2	3.9	9.2	18.9
1969-89	-9.0	-2.3	13.4	27.7	35.9	43.0

SOURCE: See Table 1. Means for the top quintile are biased downward due to top-coding.

5 MAR 1995

TABLE 3
INEQUALITY IN INCOME MEASURED IN ONE YEAR AND AVERAGED OVER FIVE YEARS

	QUINTILE			RATIO	
YEAR	BOTTOM	MIDDLE	TOP	20/50	20/80
	INCOME IN ONE YEAR				
1970	14,171	37,643	80,739	.376	.175
1975	13,777	37,998	87,165	.363	.158
1980	12,628	39,955	91,616	.316	.138
1985	9,446	40,333	93,578	.234	.101
Change: 1970-1980	-33.3	7.2	15.9	-.142	-.074
	INCOME AVERAGED OVER FIVE YEARS				
1970	16,663	38,099	78,687	.437	.212
1975	16,309	40,003	86,554	.408	.188
1980	14,798	40,168	88,184	.368	.168
1985	12,524	40,379	90,247	.309	.139
Change: 1970-1980	-24.8	6.0	14.7	-.128	-.073

Source: Tabulations by Tim Veenstra using the PSID 1989 wave. Sample includes all children.

c:\kids\irpkid.\3/5/95 (HOUSEHOLDS WEIGHTED BY KIDS)
SOURCE: VEENSTRA, INEQ.LST:9-22-94

Table 4
Income and Consumption by Income Decile or Quintile and Year: CEX Consumer Units with Children

Measure	Income decile		Income quintile			
and year	First	Second	Second	Third	Fourth	Fifth
CUs REPORTING ON INCOME FROM AT LEAST ONE MAJOR SOURCE						
Income						
1972-73	7,621	15,420	23,973	33,427	43,763	71,435
1980-81	6,510	12,298	20,941	31,399	41,501	60,929
1988-90	5,760	11,567	20,188	31,293	44,205	75,287
Consumption						
1972-73	14,082	17,926	22,529	26,814	32,756	44,482
1980-81	12,817	15,834	21,942	27,568	33,396	42,308
1988-90	13,590	16,420	20,701	26,981	34,564	49,683
Consumption as a percent of income						
1972-73	185	116	94	80	75	62
1980-81	197	129	105	88	80	69
1988-90	236	142	102	86	78	66
CUs ANSWERING ALL INCOME QUESTIONS						
Income						
1980-81	6,184	11,711	19,877	30,647	40,989	59,705
1984-86	4,280	9,709	17,999	29,031	41,223	71,473
1986-88	4,379	9,507	18,144	29,787	43,211	69,972
1988-90	5,521	10,858	18,858	29,637	41,777	70,834
Consumption						
1980-81	12,511	14,331	21,043	27,228	33,355	40,710
1984-86	15,991	14,840	19,780	26,186	31,293	46,088
1986-88	13,388	12,113	20,297	25,827	33,111	45,046
1988-90	13,132	14,938	19,910	25,469	32,307	46,862
Consumption as a percent of income						
1980-81	202	122	106	89	81	68
1984-86	374	153	110	90	76	64
1987-88	306	127	112	87	77	64
1988-90	238	138	105	86	77	66

SOURCE: Tabulations by Judith Levine and Scott Winship from data tapes prepared by John Sabelhaus. The number of consumer units with children reporting the amount of income received from at least one major source is 8,106 in 1972-73, 883 in 1980-81, 1,606 in 1984-86, 1,970 in 1987-88, and 2,794 in 1989-90.

Table 5
Consumption by Consumption Decile or Quintile, CEX Consumer Units with Children, by Year

Year	Consumption Decile		Consumption Quintile			
	First	Second	Second	Third	Fourth	Fifth
1972-73	9,858	15,104	20,004	26,000	32,834	46,688
1980-81	8,429	13,944	20,116	26,585	33,650	49,374
1985-86	8,525	13,906	19,732	26,241	34,716	53,808
1987-88	7,590	12,777	190376	26,172	34,102	52,345
1989-90	8,259	13,235	19,410	26,054	34,132	54,720
Percent change:						
1972-80	-14.5	-7.7	.6	2.3	2.5	1.4
1980-90	-2.0	-5.1	-3.5	-2.0	1.4	10.8
1972-90	-16.2	-12.4	-3.0	.2	4.0	12.4

SOURCE: See Table 4. The sample sizes are 8,131 in 1972-73, 1,173 in 1980-81, 2,355 in 1984-86, 2,570 in 1986-88, and 3,915 in 1988-90.

c:\kids\irpkid.\3/5/95
SOURCE: CEX (Weights for total sample times N of Kids).
Levine-Winship, 7OTABLE9, TABLE 9.

5 MAR 1995

Table 6
Percent of Children at Different Income Levels Living in Homes with Selected Problems: 1970 to 1990

Measure	Income decile		Income quintile				
and year	First	Second	Second	Third	Fourth	Fifth	Mean
DESIGN INADEQUACIES							
Incomplete plumbing[1]							
1970	20.5	15.5	6.6	2.4	1.9	.6	
1980	5.5	4.1	1.9	.9	.5	.1	
1990	3.2	1.3	.9	.5	.4	.3	
Change	-17.3	-14.2	-5.7	-1.9	-1.5	-.3	
Incomplete bathroom[2]							
1973-75	11.4	7.5	3.2	.9	.4	.3	2.9
1977-79	7.4	4.6	2.5	1.1	.4	.2	2.1
1981-83	6.1	4.1	2.2	1.0	.4	.2	1.8
1985-89	2.5	2.2	.8	.7	.6	.6	1.1
Change	-8.9	-5.3	-2.4	-.2	.2	.3	-1.8
No sewer or septic system							
1973-75	8.1	5.1	2.1	.6	.3	.1	2.0
1977-79	4.9	3.0	1.5	.6	.2	.1	1.2
1981-83	2.7	1.9	.9	.3	.1	0	.7
1985-89	1.7	.9	.2	.1	0	0	.3
Change	-6.4	-4.2	-1.9	-.5	-.3	-.1	-1.7
No central heat							
1973-75	46.2	42.9	30.3	18.7	12.3	6.8	22.5
1977-79	39.3	40.2	28.6	18.8	12.3	6.1	21.1
1981-83	35.7	38.1	31.9	22.2	14.7	9.1	22.9
1985-89	32.3	34.7	28.1	21.4	14.9	9.6	21.5
Change	-13.9	-8.2	-2.2	2.7	2.6	2.8	- 1.0
No electric outlets in one or more rooms							
1973-75	12.1	10.0	5.9	3.5	2.6	1.9	5.0
1977-79	8.4	6.7	5.0	2.8	1.6	1.4	3.4
1981-83	9.3	6.6	4.7	3.1	2.2	1.6	3.9
1985-89	6.0	6.0	3.8	2.4	2.0	1.1	3.1
Change	-6.1	-4.0	-2.1	-1.1	-.6	-.8	-1.9
MAINTENANCE PROBLEMS							
Holes in floor							
1973-75	8.2	5.6	2.9	1.8	.8	.6	2.6
1977-79	8.2	5.5	3.7	1.5	1.0	.6	3.4
1981-83	8.9	7.3	4.2	1.6	.8	.6	3.9
1985-89	7.0	5.8	2.6	1.4	.8	.6	3.1
Change	-1.2	.2	-.3	-.4	0	0	.5

c:\kids\irpkid.\3/5/95 (WEIGHTED BY KIDS)
SOURCE: Knutson, Newtb55.kid, 9-27-94; Veenstra, Kidwtqnt.lst, 9-29-94

5 MAR 1995

Table 6 continued

Open cracks in wall or ceiling							
1973-75	17.9	14.3	8.9	5.6	3.8	2.8	7.5
1977-79	18.5	14.4	9.4	5.0	3.5	2.5	7.4
1981-83	19.2	16.2	10.5	5.4	3.7	2.6	8.0
1985-89	19.9	15.9	10.6	6.3	4.2	3.2	8.4
Change	2.0	1.6	1.7	.7	.4	.4	.9
Leaky roof							
1973-75	16.5	14.2	9.9	7.2	5.7	5.3	8.6
1977-79	14.5	13.5	10.3	7.1	5.6	4.9	8.3
1981-83	14.9	12.8	9.9	7.0	6.0	4.9	8.3
1985-89	11.9	12.5	10.1	8.5	7.7	7.3	9.1
Change	-4.6	-1.7	.2	1.3	2.0	2.0	.5
NEIGHBORS							
Crime problem in neighborhood							
1973-75	18.9	19.1	17.1	16.5	16.4	16.6	17.1
1977-79	18.9	16.0	15.4	14.4	13.3	13.5	14.8
1981-83	19.1	18.7	15.8	14.4	14.4	14.5	15.6
1985[3]	26.3	19.6	17.0	14.1	13.3	11.8	16.0
Change	7.4	.5	-.1	-2.4	-3.1	-4.8	-1.1
CROWDING							
More than one person per room (AHS)							
1973-75	31.6	34.7	26.5	19.0	15.6	11.6	21.2
1977-79	26.1	28.5	22.1	14.9	11.1	8.5	16.8
1981-83	22.7	26.7	21.0	13.5	8.0	5.9	14.6
1985-89[4]	19.2	23.4	17.6	10.9	7.3	5.3	12.5
Change	-12.4	-11.3	-8.9	-8.1	-8.3	-6.3	-8.7
OWNERSHIP							
Tenant (AHS)							
1973-75	62.5	54.5	38.3	23.9	15.2	9.1	29.1
1977-79	67.0	58.8	39.6	21.6	12.7	7.2	28.8
1981-83	67.8	62.2	44.6	24.8	14.6	7.6	31.4
1985-89	78.2	68.9	50.0	31.0	18.0	8.1	36.2
Change	13.7	14.4	11.7	7.1	2.8	-1.0	7.1

SOURCES: Measures shown for 1970, 1980, and 1990 are from the decennial Census (tabulations by David Knutson), while those shown for 1973 through 1989 are from the AHS (tabulations by Tim Veenstra). In the Census, the unweighted sample sizes for the bottom decile are between 2,700 and 3,500. In the AHS they are 7,638 in 1973-75, 5,033 in 1977-79, 4,424 in 1981-83, and 4,027 in 1985-89. The AHS income data are for families rather than households.

1. Hot and cold water, sink, toilet, and shower or tub for the exclusive use of household members. Plumbing facilities need not be inside respondent's apartment in 1970, but must be in the building.
2. Complete plumbing located in a single room within the unit.
3. Respondent's judgment. Data not available after 1985.
4. Room count increased slightly in 1985 due to questionnaire change.

5 MAR 1995

Table 7
Percentage Point Gap between Lowest Decile and Middle Quintile of Households with Children, by Year

	Percentage Point Gap			Change in Gap[a]	
Decennial Census	1970	1980	1990	1970s	1980s
Incomplete plumbing	18.1	3.6	2.3	-	o
American Housing Survey	1973-75	1977-79	1985-89		
Crime problem in neighborhood	2.4	4.5	12.2	+	+
Holes in floor	6.4	6.7	5.6	o	o
No sewer or septic system	7.5	4.3	1.6	-	-
No electric outlet in one or more rooms	8.6	5.6	3.6	-	-
Roof leaks	9.3	6.6	3.4	-	-
Incomplete bathroom	10.5	6.3	1.8	-	-
Open cracks in wall or ceiling	12.3	13.5	13.6	o	o
More than one person per room	12.6	11.2	8.3	o	-
No central heat	27.5	20.5	10.9	-	-
Tenant	38.6	45.4	47.2	+	o

SOURCE: Arithmetic difference between the percentages for the middle quintile and the bottom decile in Table 3.

a. + designates an increase in the gap of two or more percentage points.
- designates a decrease in the gap of two or more percentage points.
o designates a change of less than two percentage points.

5 MAR 1995

Table 8
Effects of Family Income on Children's Annual Doctor Visits, by Year

Children's age and Year	Mean	Regression coefficient of ln family income[a] B	(SE)	Income of bottom decile as proportion of median[b]	Estimated visits for children in bottom decile[c]
Percent with a Doctor Visit Last Year					
Under 7					
1970	82.1	.094	(.005)	.176	65.8
1980	87.6	.028	(.004)	.138	82.1
1982	88.2	.033	(.004)	.137	81.1
1989	90.0	.037	(.003)	.123	82.3
7 to 17					
1970	63.0	.099	(.004)	.176	45.8
1980	69.3	.044	(.004)	.138	60.6
1982	70.1	.047	(.005)	.137	60.8
1989	73.8	.061	(.004)	.123	61.0
Number of Doctor Visits in a Year					
Under 7					
1970	4.8	.393	(.150)	.172	4.1
1980	4.7	.226	(.098)	.138	4.3
1982	4.2	.267	(.062)	.137	3.7
1989	4.4	.351	(.072)	.123	3.7
7 to 17					
1970	3.4	.087	(.081)	.172	3.3
1980	3.3	.032	(.068)	.138	3.2
1982	3.3	.055	(.082)	.137	3.2
1989	3.5	.073	(.080)	.123	3.4

SOURCE: Tabulations by David Knutson from HIS public use data tapes. Sample sizes range from 10,000 to 14,000 for children under seven and from 16,000 to 25,000 for children aged 7 to 17.

a. Controlling ln family size.

b. Estimated from data on all children under the age of 18 in 1969, 1979, and 1989 (see Table 2), assuming no change in 1969-70 and 1979-80, and by linear interpolation for 1982.

c. Column 1 - (Column 2)(ln[Column 4]). The median closely approximates the geometric mean for all households with children. The arithmetic mean for the bottom decile, used to calculate column 5, exceeds the geometric mean but is probably a better measure of this group's true income.

c:\kids\irpkid.\3/5/95 (WEIGHTED BY KIDS)

Table 9
Percent of Children at Different Income Levels with Selected Consumer Durables and Telephone Service: 1970 to 1990

Measure	Income decile		Income quintile			
and year	First	Second	Second	Third	Fourth	Fifth
Motor vehicle (AHS)						
1973-75	62.6	80.5	91.6	97.3	98.5	99.2
1977-79	61.5	80.2	92.2	98.1	99.3	99.7
1981-83	63.9	76.6	91.9	97.9	99.2	99.5
1985-89	56.8	77.5	92.7	97.8	99.0	99.3
Change	-5.8	-3.0	-.9	.5	.5	.1
Motor vehicle (Census)						
1970	59.8	76.4	90.4	95.6	97.6	98.8
1980	58.6	78.1	89.7	95.7	97.7	98.4
1990	57.3	82.1	91.7	97.0	98.0	99.0
Change	-2.5	5.7	.7	1.4	.4	.2
Two or more vehicles (Census)						
1970	13.2	20.0	32.3	44.4	57.6	74.8
1980	14.2	21.0	35.3	50.7	64.7	76.6
1990	17.3	34.3	56.4	75.3	86.6	92.9
Change	4.1	14.3	24.1	30.9	29.0	18.1
Air conditioning (AHS)						
1973-75	27.5	31.8	41.1	48.9	55.2	62.2
1977-79	30.9	33.6	45.2	53.1	58.3	65.1
1981-83	36.6	39.6	49.1	57.3	63.7	69.2
1985-89	41.5	47.4	57.9	64.9	69.7	72.8
Change	14.0	15.6	16.8	16.0	14.5	10.6
Clothes washer (CEX)						
1972-73	62.8	72.8	84.2	91.5	95.3	96.3
1984-89	57.8	61.4	78.6	84.4	92.8	97.1
Change	-5.0	-11.4	-5.6	-7.1	-2.5	.8
Clothes dryer (CEX)						
1972-73	23.3	38.3	59.6	73.9	83.1	91.0
1984-89	37.5	38.0	62.0	75.2	88.9	94.6
Change	14.2	-.3	2.4	1.3	5.8	3.6
Dishwasher (CEX)						
1972-73	9.1	10.1	18.0	31.0	45.5	68.7
1984-89	16.5	16.0	25.8	41.6	58.2	79.7
Change:						
1972-90	7.4	5.9	7.8	10.6	12.7	11.0

continued next page

Table 9 continued

Measure and year	Income decile		Income quintile			
	First	Second	Second	Third	Fourth	Fifth
Telephone (Census)						
1970	60.8	66.9	83.0	91.7	95.0	98.5
1980	72.1	80.2	88.7	95.8	98.3	99.0
1990	68.7	79.7	90.8	96.5	98.3	99.5
Change	7.9	12.8	7.8	4.8	3.3	1.1

SOURCES: For Census and AHS data see Table 6. Data on clothes washers, clothes dryers, and dishwashers are from the Consumer Expenditure Survey (tabulations by Judith Levine and Scott Winship using tapes prepared by John Sabelhaus). The unweighted sample sizes for the bottom decile in the CEX are roughly 800 in 1972-73 and 640 in 1984-89. The CEX income data are for the consumer unit.

c:\kids\irpkid.\3/5/95
SOURCE: CEX (Weights for total sample times N of Kids).
Levine-Winship, 70TABLE9, TABLE 9.

5 MAR 1995

TABLE 10
Labor Force Participation and Employment among Parents

	Married Fathers	Married Mothers	Single Fathers	Single Mother
	Labor force Participation			
1970	93.9	37.5	81.9	51.3
1974	93.2	40.8	82.5	51.4
1976	92.6	44.1	80.6	54.9
1980	92.9	51.5	85.2	59.9
1985	92.6	57.6	85.3	59.0
1987	92.3	60.6	87.3	61.8
1990	92.3	62.9	83.9	61.9
1992	92.1	64.6	84.3	60.1
Change				
1970-80	-1.0	14.0	3.3	8.6
1980-90	-1.6	13.1	.9	.2
	Unemployed			
1970	2.6	6.2	1.1	9.2
1974	3.0	5.6	6.0	8.8
1976	5.2	8.0	8.0	12.9
1980	4.7	6.0	10.1	12.2
1985	5.4	6.8	11.7	14.3
1987	5.0	5.6	9.7	13.6
1990	3.8	4.3	7.9	11.0
1992	6.2	5.5	9.5	13.4
Change				
1970-80	2.1	-.2	9.0	3.0
1980-90	.9	-.5	- .6	1.2
	Employed			
1970	91.5	35.2	80.9	46.7
1974	90.4	38.5	77.5	46.9
1976	87.7	40.6	74.1	47.9
1980	88.6	48.4	76.6	52.6
1985	87.6	53.7	75.3	50.6
1987	87.7	57.2	78.8	53.4
1990	88.7	60.2	77.3	55.1
1992	86.4	61.1	76.3	52.1
Change				
1970-80	-2.9	13.2	-4.3	-5.9
1980-90	-2.2	11.8	-.3	-.5

Source: Tabulations by David Knutson using data from March CPS file described in the Appendix.

c:\kids\irpkid.\3/5/95 (WEIGHTED BY KIDS IN UNIT)
SOURCE: Knutson

10-2-94

5 MAR 1995

Table 11
Hours Parents Worked Last Week, by Year

All	Married Fathers	Married Mothers	Single Fathers	Single Mothers
1970	40.4	10.8	34.1	15.4
1974	39.4	11.8	31.5	16.1
1976	37.7	12.4	29.8	16.8
1980	38.0	15.3	31.2	18.8
1985	38.3	17.1	30.8	17.8
1987	38.3	18.4	32.1	19.1
1990	39.0	19.5	31.8	19.9
1992	38.1	20.3	30.8	18.6
Change				
1970-80	-2.4	4.5	-1.9	3.7
1980-90	1.0	5.2	-.4	-.2

Workers	Married Fathers	Married Mothers	Single Fathers	Single Mothers
1970	45.5	32.1	43.1	35.3
1974	45.2	32.3	42.7	36.0
1976	44.5	32.2	41.8	37.1
1980	44.5	32.6	42.5	37.2
1985	45.1	33.1	42.5	37.1
1987	45.0	33.6	42.7	37.2
1990	45.4	33.8	43.0	37.9
1992	45.3	34.6	41.5	37.2
Change				
1970-80	-1.0	.5	-.6	1.9
1980-90	.8	1.2	.5	.7

Source: See Table 10.

c:\kids\irpkid.\3/5/95 (WEIGHTED BY KIDS IN UNIT)
SOURCE: Knutson

10-2-94

5 MAR 1995

Table 12
Employment and Work Status of 0 to 3 Year Old Children's Parents

	Employed			
	Married Fathers	Married Mothers	Single Fathers	Single Mothers
1970	90.2	24.3	79.1	36.5
1974	89.7	28.0	64.4	35.8
1976	86.3	29.0	72.3	36.7
1980	87.9	37.7	68.0	35.8
1985	87.1	45.3	72.2	35.3
1987	87.7	49.9	79.7	41.6
1990	89.1	52.1	72.6	41.1
1992	85.5	53.2	73.2	40.4
Change				
1970-80	-2.3	13.4	-11.1	-.7
1980-90	1.2	14.4	4.6	4.4

	Workers' Hours			
	Married Fathers	Married Mothers	Single Fathers	Single Mothers
1970	45.0	31.6	44.5	34.9
1974	44.9	31.7	43.2	36.7
1976	44.0	31.6	41.2	36.6
1980	44.4	31.2	41.2	36.5
1985	44.9	31.9	40.7	36.7
1987	44.9	32.7	42.9	36.4
1990	45.3	32.9	41.1	36.7
1992	45.5	33.6	41.4	35.5
Change				
1970-80	-.6	-.4	-3.3	1.6
1980-90	.3	1.7	-.1	.2

Source: See Table 10.

c:\kids\irpkid.\3/5/95 (WEIGHTED BY KIDS IN UNIT)
SOURCE: Knutson

10-2-94

5 MAR 1995

Table 13
Potential hours per week available for children

	Per Child's Family	Per Child
	All Children	
1970	162.8	68.1
1974	159.9	72.0
1976	159.7	74.2
1980	154.0	77.5
1985	150.3	78.3
1987	148.7	78.3
1990	146.7	76.4
1992	145.6	76.0
Change 1970-92	-17.2	7.9
	Children Less Than 3 Years Old	
1970	168.7	85.0
1974	166.3	92.0
1976	166.8	93.8
1980	161.7	93.5
1985	157.0	90.3
1987	154.9	89.6
1990	151.9	85.7
1992	150.9	85.9
Change 1970-92	-17.8	.9

Source: See Table 10.

c:\kids\irpkid.\3/5/95 (WEIGHTED BY KIDS IN UNIT)
SOURCE: Knutson

10-2-94

5 MAR 1995

Table 14
Annual Hours of Housework and Market Work For Families with Children, 1976 to 1987.

Year	House-Work	Market Work	Total Hours
1976	1,778	2,534	4,312
1977	1,864	2,568	4,433
1978	1,853	2,606	4,459
1979	1,793	2,672	4,465
1980	1,808	2,639	4,447
1981	1,773	2,623	4,396
1983	1,702	2,500	4,202
1984	1,642	2,571	4,213
1985	1,589	2,706	4,295
1986	1,540	2,689	4,229
1987	1,514	2,693	4,207
Change 1976-87	-264	159	-105

Source: Tabulations by Tim Veenstra using the 1989 wave of the PSID. The sample is all households with children.

c:\kids\irpkid.\3/5/95 (WEIGHTED BY KIDS IN UNIT)
SOURCE: Knutson

10-2-94

Longitudinal Indicators of Children's Poverty and Dependence

Greg J. Duncan
Center for Urban Affairs and Policy Research
Northwestern University
2040 Sheridan Road
Evanston, IL 60208
Telephone: 708-467-1503
E-mail: greg-duncan@nwu.edu

April 20, 1995

The work for this paper was supported in part by the National Institute for Child Health and Human Development as part of its Family and Child Well-being Research Network. Many of the ideas for indicators based on data from the Survey of Income and Program Participation were developed while the author participated on the National Research Council's Panel to Evaluate the Survey of Income and Program Participation. Pat Ruggles and Connie Citro made extensive contributions to the social-indicators portion of the Panel's work. This paper has benefitted from helpful comments from Terry Adams, Rebecca Blank, Jeanne Brooks-Gunn, Brett Brown, Sheldon Danziger, Dorothy Duncan, Pamela Klebanov and Kristin Moore.

Longitudinal Indicators of Children's Poverty and Dependence

I. Introduction

Monthly measures of unemployment and consumer-price inflation plus quarterly reports on aggregate disposable income are the best-known social indicators of household-sector well-being in most Western countries. Unique to the United States is the production of well-publicized annual reports on the extent of poverty among various groups, including children. The singular position of the United States in the routine compilation of poverty statistics results from a number of factors, including: i) a basic consensus that a comparison of a household's total income and its family-size-based "official" poverty threshold says something meaningful about whether individuals living in that household have a minimum level of material resources; ii) a national statistical office brave enough to ask questions about income of a nationally-representative sample, coupled with a population willing and able to provide reasonably accurate responses to such questions; and iii) a national psyche willing to absorb periodic reports of poverty indicators and eager to debate the policy implications of the statistics. These factors, combined with periodic reports on welfare receipt, health-insurance coverage and related topics, produce a comparatively rich set of indicators of child-based economic deprivation in the United States.

Unfortunately, extensive research on the nature and consequences of economic deprivation in the United States yields many reasons to be dissatisfied with the current set of indicators of children's deprivation and dependence. A National Research Council committee is about to propose a new method for measuring income-based poverty. Data from the Survey of Income and Program Participation (SIPP) suggest that the survey that is used to produce poverty indicators -- the Current Population Survey (CPS) -- badly undercounts annual income and therefore overcounts the poor (e.g., U.S. Bureau of the Census, 1991).[1] Research on expenditure-based indicators of deprivation reveals many troubling cross-sectional and time-series inconsistencies with income-based indicators (Mayer, 1994).

The situation with respect to social indicators of welfare receipt or dependence is worse yet, since, despite well-publicized calls for such indicators (e.g., by Moynihan in The New York Times, 1991), there are no routinely-released indicators of welfare use.

This paper takes a longitudinal perspective in assembling its list of sources of dissatisfaction with current indicators of poverty and dependence and in making recommendations for change. It begins by outlining the problem of describing dynamic processes such as economic deprivation, welfare use or unemployment experiences using either longitudinal or cross-sectional data. It then presents examples of longitudinal indicators for which timely data do not usually exist. It concludes with specific recommendations for indicators that could be produced with available data.

II. Describing dynamic processes

The task of describing dynamic processes with any type of data, whether cross-sectional or longitudinal, is formidable. Figure 1 displays ten different possible patterns of "economic deprivation" over the twenty-two-year period from 1979 to 2001. The mixture of short- and long-term patterns is chosen to be roughly consistent with findings from the literature on periods of receipt of benefits from the Aid to Families with Dependent Children transfer program; the essential features of these patterns, however, are similar to those found in connection with other aspects of economic deprivation such as poverty or, if the time scale were more compressed, unemployment.

The line labeled "1" depicts a period of continuous receipt over the entire 22-year period. Individual "2" has a lengthy period of receipt that is divided into two spells, the first running from 1982 to 1988 and the second running for three years beginning in 1992. Individuals 4, 5 and 6 have only short, single spells, while the last four individuals have diverse experiences that could be described as intermediate in total length.

Important to note from these patterns are the following features:

- Experiences are extremely heterogeneous, with substantial fractions of individuals' deprivation experiences lasting no more than 2 years and equally substantial fractions lasting for quite long periods.[2]

- Repeated episodes are common, occurring in close to half of the cases and sometimes at wide intervals.[3] Thus an analysis of individual spells (e.g., the first spells of individuals 3 and 10) can provide a badly biased picture of the total scope of an individual's longer-run experience with deprivation or dependence.

- The total length of deprivation or dependence can be decomposed into several components: i) the incidence of a first spell; ii) the duration of the first spell; iii) the spacing and length of second and subsequent spells (Ellwood, 1986; Gottschalk and Moffitt, 1994).

Patterns depicted in Figure 1 have not been linked to the demographic or other characteristics of the individuals. In the case of children, "childhood" or "early childhood" define important periods over which patterns of deprivation or dependence might be measured. In terms of Figure 1, this amounts to superimposing a fixed-length window corresponding to the childhood period of interest. In the case of, say, a six-year window from birth to a child's sixth birthday for children born at the beginning of 1983, this takes the form of the six-year shaded portion of Figure 1. In the case of individual 2, the birth occurs one year after the beginning of a spell of poverty or dependence for the mother. For individuals 1 and 9, the period of deprivation or dependence extends beyond the sixth birthday. And, in the case of all individuals other than 4 and 5, the six-year window misses episodes of deprivation or dependence occurring later in childhood.

Not shown in Figure 1 but also noteworthy is the importance of subannual detail on many experiences, particularly short ones.[4]

- Precise description of short-run experiences with poverty or welfare receipt usually require data collected over subannual accounting periods.

These various features have important implications for analyzing experiences with poverty and welfare use (Bane and Ellwood, 1983):

- Data describing the mixture of short and long-term experiences will differ dramatically depending on how the experiences are sampled. Taking welfare receipt as an example, the ten individuals depicted in Figure 1 constitute a group of individuals who ever received welfare during the 22-year period (an "ever-on" sample.) Equal fractions (30% in this case) of "ever-on" recipients have short-term and long-term patterns. In contrast, a sample drawn at any given point (a "point-in-time" sample) will have far greater concentrations of long-term than short-term recipients. (In 1994, for example, 50% of the 10 individuals are recipients and none of them are short-term.) The reason for this difference is clear: long-term recipients have a much greater chance than short-term recipients of showing up at any given point in the 22-year period.

- Differences between "ever-on" and "point-in-time" samples speak to different policy concerns. The "ever-on" sample describes the distribution of experiences of all individuals who ever come into contact with the system and is important for thinking about policies such as time limits that might be instituted for all new recipients. The point-in-time sample describes how the benefits are distributed and the nature of a group affected by policies directed at the caseload at any given point. Thus, accurate descriptions of both "ever-on" and "point-in-time" samples are essential.

Deprivation and dependence have both intra- and inter-generational dimensions. If the patterns of deprivation or dependence shown in Figure 1 span both childhood and adulthood, then they can be taken to show both kind of patterns. Suppose, for example, that all of the individuals depicted in Figure 1 became adults (in the sense of forming their own households and/or being at risk for receiving benefits on behalf of their own children) at the beginning of 1995. Deprivation or dependence spells prior to the beginning of 1995 refer to their parental households, while deprivation or dependence after 1995 reflect their own experiences as adults. Of the five individuals leaving poor or dependent families, two (numbers 1 and 3) continue to be poor or dependent, while one individual (number 6) who had not been poor or dependent during childhood became so shortly after reaching adulthood. These patterns are not inconsistent with the intergenerational literature, which points to heterogeneous experiences, with positive but far from perfect correlations in economic status or welfare use across generations.[5]

The importance of the length of the accounting period and observation window. The ability of surveys and administrative data to describe patterns of deprivation or dependence is governed by the accounting periods over which such experiences are measured and the total length of their observation windows. In terms of patterns depicted in Figure 1, the ideal data for describing welfare experiences would be month-by-month observations on individuals over the entire 22-year period. The monthly detail would capture the short-run dynamics of deprivation, program eligibility and dependence, while the 22-year coverage would provide information on multiple spells and minimize problems with observations being censored by the beginning or end of the observation window.

Going from optimal to second-best data is not unproblematical, given the heterogeneity of experiences and competing demands for information about short and longer-run dynamics. Indicators of short-run dynamics require data with a subannual accounting period. Even if collected over an observation window as short as one or two years, such data would be valuable for describing rates of transitions into and out of deprivation or dependence as well as events (e.g., marital, employment-related) associated with those transitions. For welfare use it might be possible to gather reliable retrospective information about the duration of receipt or nonreceipt prior to the beginning of the survey period.[6] This would enable analysts to classify spells observed during the observation window as first or subsequent spells, as well as to determine the length of censored spells.

Accurate description of longer-run dynamics requires a longer observation window, and not so much monthly detail. Gottschalk and Moffitt (1994) argue for the utility of "total time on" and "total fraction of income" measures of welfare use, in which the total number of years of welfare use and percentage of total income made up by welfare payments are calculated over a multi- (in their case, seven-) year observation period, without regard to the particular pattern of spells. Duncan et al. (1984) and Duncan and Rodgers (1991) develop analogous measures for poverty.

The picture drawn by an observation window of, say, ten years, can be seen in Figure 1 by taking the patterns observed between the lines drawn at the beginning of 1980 and 1990. Nine of the ten individuals whose experiences are depicted in Figure 1 are caught by this ten-year window. There appear to be three short-term patterns of two years or less (individuals 4, 5 and 10) and three long-term patterns of five years or more (individuals 1, 2 and 7), which is not substantially different from the distribution of 22-year patterns. Only one individual (number 10) is seriously misclassified in the sense that his second and longer spell is missed by the 1980-1990 window. Individual 10 *is* correctly classified if the observation window is taken literally -- i.e., his experiences during the decade of the 1980s were short-term.

Ellwood's analysis of long-term welfare experiences takes a different approach, in which the incidence and duration of first spells and the timing and length of subsequent spells are combined into a simulation model of total lifetime welfare experiences. Data requirements for this approach differ little from those of the "total time on" approach. In both cases one needs a long enough observation window, possibly supplemented with retrospective data, to identify first spells and gauge the length and distribution of first and subsequent spells.

III. What dimensions of economic deprivation are most important?

Given that actual patterns of deprivation and dependence include both short-term and long-term experiences, it is useful to step back and ask: Under what circumstances should social policy and therefore social indicators attach importance to the duration of deprivation or dependence?

There are two ways of approaching this question. The first is to look at the extent to which current or contemplated policies take into account duration in their program rules. The distribution of effort in developing short- and longer-run indicators of deprivation and dependence should bear some correspondence to the distribution of short and longer-run definitions of deprivation found in actual programs. If, for example, most programs opt for a monthly income-accounting period, then it would make sense from this perspective to develop at least some social indicators based on monthly data. If welfare reform imposes a 24-month limit on the duration of receipt of welfare, then there is an obvious need for monthly information about welfare experiences that span periods of more than 24 months.

The second approach to thinking about the important features of duration of deprivation is to ask what difference duration and timing of deprivation make for children's development. If, in contrast to the situation for longer-run poverty or dependence, there are no discernible detrimental effects of short-run poverty or welfare dependence on children's IQs or academic achievement or on their welfare or labor-supply behavior when they become young adults, then social indicators should be less concerned with short-run episodes of deprivation or dependence than with longer-term experiences. And if the evidence indicates that poverty or welfare receipt affects development during early childhood but not during adolescence, then social indicators of childhood poverty and dependence should be especially concerned with descriptions of patterns during early childhood.

Program design. With respect to the question of actual practice, it is clear that many social-assistance programs are aimed at fulfilling short-term needs -- food or heating for example -- and that almost all means-tested programs in the United States rely on a monthly accounting period for the allocation of their benefits. Since programs do not take into account whether families with little incomes and few assets have had or will soon again have adequate levels of economic resources, at least some social indicators should be similarly unconcerned with the longer-term picture. Thus:

- The policy importance of short-term needs dictates that at least some social indicators of deprivation and dependence focus on "point-in-time" samples, monthly accounting periods and short-term dynamics experiences.

Policy initiatives focused on curing long-term poverty or preventing long-term welfare dependence must make the distinction between the short and longer term, recognizing which poor people are most likely to remain poor as well as which of the long-term poor would profit most from these programs. Ellwood (1986) argues that it is most effective to target training programs designed to promote work-to-welfare transitions on employable "would-be" long-term poor or social-assistance recipients. His accounting period for lifetime welfare use--25 years--is long indeed. Thus:

- Many policy issues focus on long-term poor or dependent families, dictating a need for long-run indicators.

Developmental consequences. A different perspective is provided by evidence on the developmental consequences of short- and long-term deprivation and dependence. Does it really matter for children's development whether their childhood episodes of economic deprivation are short- or longer-term? Are a few years of deprivation sufficient to leave developmental scars or is the longer-run level of resources of primary importance?

It seems reasonable to expect that being poor for relatively short periods is less detrimental to children than are sustained bouts of poverty. At the same time, if families move above the poverty line, but not very far above it, then the duration of poverty may make little difference since their income has not risen enough to enable families to make the changes--e.g., moving to a better neighborhood, purchasing high-quality childcare, investing in a beneficial home-learning environment--that would produce measurable improvements in their children's development.

What little evidence there is suggests that duration does indeed matter. Miller and Korenman (1994) use mother-child data from the National Longitudinal Survey of Youth to test for effects of family income on the likelihood of "stunting" (low height for weight) and "wasting" (low weight for height) among young children. They find that a measure of income in the year prior to the measurement of physical characteristics is a much less powerful predictor of physical-health problems than is a measure of income averaged over the ten years prior to the measurement of physical characteristics. Corcoran et al. (1992) find that the number of years adolescents lived in families with incomes below the poverty line was a highly significant predictor of school attainment and early career outcomes, even after controlling for average level of family income. Duncan et al. (1994) show that age-5 IQs of a sample of low-birthweight children were significantly lower if those children had spent all as opposed to part of their childhood living in families with incomes below the poverty line, even after controlling for more conventional measures of socioeconomic status such as maternal schooling and family structure. And the IQs of children living in households with income consistently above the poverty line were significantly higher than the IQs of the part-time poor children.

Timing of poverty may also influence development, although there is scant evidence on this issue. Haveman et al. (1991) use nationally-representative data spanning 20 years and find that the combination of poverty and welfare use between ages 12 and 15 is a significant predictor of high-school dropout status, whereas combined poverty and welfare use at earlier periods in childhood is not. Duncan et al. (1994) find no effect of the timing of poverty in early childhood on either age-5 IQ or behavioral problems.

Similar questions can be asked of the literature on welfare receipt: does the length of parental dependence matter for children's outcomes? Theories of poverty have often included an intergenerational component, with anthropological studies arguing that children growing up in poor families and communities are likely to adopt the fatalistic and self-defeating attitudes and behaviors of their parents (Lewis, 1968). In the case of welfare use, it is also easy to imagine that characteristics of some recipient households -- lack of attachment to the labor force and dependence on government income support -- might convey to children the viability of similar kinds of lives in adulthood.

An obvious problem in drawing conclusions about the intergenerational consequences of parental welfare receipt is the need to adjust for other aspects of parental background and environment that may also affect a child's chance of subsequent success. Children from AFDC-dependent homes generally have fewer parental resources available to them, live in worse neighborhoods, go to lower-quality schools, and so forth. Any of these factors could have an effect on their accomplishments as adults that is independent of their parents' AFDC receipt. While the most recent literature appears to indicate that parental welfare receipt does indeed matter, it provides mixed evidence on the importance of duration and timing.

Corcoran et al. (1992) find that parental welfare income is a negative and statistically significant predictor of the annual earnings, wage rates, work hours and family incomes of young-adult men. An et al. (1993) find no significant effect of parental receipt on the likelihood of a daughter giving birth, but a marginally significant effect of parental welfare receipt on the daughter's receipt, conditional on the daughter having a teen out-of-wedlock birth. For a sample of black children, Guo, Brooks-Gunn and Harris (1992) relate risk of repeating a grade prior to high school to family-level poverty and welfare experiences. They find that while welfare receipt immediately before the time at which the risk of grade failure is assessed becomes insignificant in the presence of controls for family SES, longer-run measures of welfare receipt remain significant even when these control variables are included in the analysis. Duncan and Yeung (1995) find similar detrimental effects of parental welfare receipt on the completed schooling of children. Gottschalk (1994) estimates models that attempt to purge the parental welfare measure of its sources of noncausal correlations with the outcomes of interest and finds highly significant effects of parental welfare receipt on the chances that daughters will have AFDC-related births. Furthermore, the strongest effects are for parental receipt immediately prior to the daughter's possible fertility.

With respect to the effects of the timing of welfare receipt, a twenty-year prospective study of over 300 urban black families in which a teenage birth had occurred in the late 1960s showed that receiving AFDC in the young childhood years had a greater effect on educational attainment (grade failure and literacy at age 19) than did welfare receipt in the young adolescent years (Furstenberg et al., 1987; Baydar, Brooks-Gunn, & Furstenberg, in press). These same studies showed that family welfare status was highly predictive of teenage pregnancy, although it was not associated with levels of academic functioning and achievement (Furstenberg, Levine, & Brooks-Gunn, 1990).

All in all, program considerations suggest a need for both short- and longer-run indicators of deprivation and dependence. There appears to be a growing consensus that both parental poverty and welfare use can have measurable effects on children's development. Since there is insufficient evidence on the impact of the duration or timing of poverty or welfare use during childhood, both short- and longer-term indicators are clearly needed.

III. Are current indicators adequate?

Thus armed with an appreciation of the utility of both short and longer-run indicators of deprivation and dependence, we turn to an assessment of our current stock of indicators.

The Current Population Survey. Most indicators of poverty come from the Census Bureau's Current Population Survey and are published in the annual volumes entitled Poverty in the United States (e.g., U.S. Bureau of the Census, 1993). Each March CPS measures income and poverty thresholds over a single, annual accounting period. The poverty status of all individuals and households in the 60,000-household CPS sample is determined and then tabulated according to a myriad of demographic characteristics.

Recent years have seen numerous attempts to gauge the sensitivity of "official" poverty estimates to the method of inflation adjustment; the inclusion of noncash sources of income such as Food Stamps and Medicaid benefits; the proration of the poverty threshold to the composition of the family during the calendar year in which income was received; and so on. When the annual CPS data are placed side by side, they form a useful time series of snapshot pictures of the incidence of annual poverty dating back to the mid-1960s. These annual poverty indicators are released at the same time each year, amid great publicity, and often generate a productive discussion in editorials, opinion-page columns and television reports.

When judged against our criteria for desirable properties of indicators of deprivation, how well does the CPS measure stack up? Unfortunately, not very well at all. In fact, were we starting from scratch in developing social indicators of economic deprivation, it would be hard to imagine selecting a worse indicator than one based on the CPS and an annual accounting period.

A first problem is the serious underreporting of transfer income in the CPS. U.S. Bureau of the Census (1993, Table C-1) reports that the CPS accounts for only 71.6% of AFDC benefits, 89.0% of Supplemental Security Income and 86.2% of other public assistance. As mentioned earlier, the CPS poverty rate is 30% higher than measured in the higher-quality SIPP data. It is puzzling that a 30% bias is not viewed with more concern than seems to be generated by the CPS bias.

A second problem with the CPS poverty measure is its annual accounting period, an example of which is depicted in Figure 1 for the calendar year 1994. The window captures as poor half of the ten individuals who were ever poor over the 22-year period, but its "point-in-time" nature leads it to miss short-term recipients altogether. Thus, a 12-month accounting period cannot be used to describe the distribution of experiences for the "ever-on" deprived or dependent.

More generally, a 12-month accounting period is not ideal as either a short- or long-run poverty indicator. It is not short enough to capture month-to-month dynamics important for program participation; nor is it long enough to capture the essential features of "long-term" deprivation. Nor can one argue that an annual accounting period is a useful "compromise" between needs for short- and longer-run periods. Since heterogeneity is the essential empirical feature of both patterns of deprivation and program needs, it is crucial to have measures of both long- and short-run deprivation rather than a compromise that fails to capture the essential features of either.

Since the CPS does not (and, given memory problems, cannot) ask about poverty for any year prior to the calendar year just preceding the March interview, nor does it provide data on intra-year (e.g., monthly) income dynamics, its annual accounting period is also ill-suited for describing trends in any of the key components of patterns of poverty and dependence: onset and duration of initial spells and spacing and length of second and subsequent spells. The absence of subannual or multi-year data also renders it incapable of describing events associated with the beginning or ending of spells. To be fair, we should note that the CPS was designed to be a labor-force survey and its annual income information has the status of a "supplement" that is administered in only one month (March) of the year.

The Survey of Income and Program Participation. The SIPP was begun largely in response to the limitations of the CPS in providing needed details on income dynamics and program participation. SIPP's panels have varied in size from 13,000 to 21,500 households. The 1993 panel has 20,000 households. SIPP's observation window is wider than that of the CPS and its thrice-a-year interviews provide data over a monthly accounting period. Panels begun between 1984 and 1995 were designed to run for 2.5 years. Beginning in 1996, the Census plans to change the sample design and field non-overlapping panels of 50,000 households, to be followed for a total of 52 months. In addition, a special SIPP panel, the Survey of Program Dynamics, or SPD, is being designed to last for ten years. The SPD will make special efforts to collect data on the children in SIPP households. It will be designed as an extension of the SIPP panel begun in 1993.

The core SIPP questionnaire, repeated every four months, asks detailed questions concerning employment, income, and participation in federal social-support programs. Much of the information is collected on a month-by-month basis. Questions are asked about all adults age 15 and over in the household. Special modules covering personal history and data on school enrollment and financing are administered once or twice to each panel.

In addition, there are a number of special topical modules. Some have been asked of every panel to date; others have been fielded only once or twice. Topics include child-care arrangements, child-support agreements, functional limitations and disability, utilization of health-care services, support for non-household members, and others.

SIPP's features are better suited to the task of describing the dynamics of deprivation and dependence, but some problems remain. An examination of Figure 1 shows that a 52-month accounting period is much more likely to capture a mixture of short- and long-term recipients, although it is still a biased sampling of the "ever-on" population. Complete spells lasting more than 52 months will not be observed in their entirety in SIPP, nor will repeat spells that are spaced more than 52 months apart. A serious problem for longitudinal indicators of deprivation and dependence is that current plans for nonoverlapping samples in SIPP introduce a very unhelpful break in SIPP-based time series on many potential dynamic social indicators. For example, it would be helpful to use data from adjacent years to calculate rates of transition out of and into poverty among children. Nonoverlapping samples between years t and t+1 render it impossible to compute transition rates between those years.

On the plus side, however, the 52-month panel period is sufficient to observe many transitions into and out of poverty and onto and off welfare rolls, as well as providing ancillary information needed to couple these transitions with events such as marriage/divorce and employment/job loss. Monthly data from the Survey of Income and Program Participation have been used to provide a number of interesting indicators of poverty and welfare incidence and transitions (e.g., U.S. Bureau of the Census, 1991 and 1992).

Welfare caseload statistics. Apart from recent SIPP-based reports on receipt of benefits from various transfer programs, the most comprehensive source of time-series information on "point-in-time" welfare samples is caseload data presented periodically in the "Green Book" of the House Ways and Means Committee. For the AFDC program, for example, the Green Book provides useful information on: i) total spending on AFDC benefits, ii) state benefit levels; iii) number of child recipients; iv) demographic characteristics of recipient families; v) past duration of receipt; and vi) fraction of recipients with no reported income other than AFDC.

Longer-run surveys. Although not used to report "official" statistics, the Panel Study of Income Dynamics (PSID) and National Longitudinal Survey of Youth (NLSY) have provided a wealth of longer-run intra- and intergenerational data on both deprivation and dependence. The PSID began with a representative sample of households in 1968 and provides annual data on income and, since 1983 for certain transfer incomes such as AFDC, monthly data on dependence for its sample households. By following children as they leave home and counting new births as part of its sample of individuals, the PSID has a mechanism for providing continuously-representative household samples (except for immigration) throughout its life as well as representative intergenerational data.

The National Longitudinal Survey was begun in 1979 with a nationally-representative sample of 14-21 year olds. It has taken annual interviews with its sample since 1979 and conducted extensive assessments of the children of the mothers in the cohort every two years beginning in 1986. Interviews taken with parents of members of the original cohorts provide rich intergenerational information. Extensive cognitive and behavioral information on children born to women in the original cohorts has been gathered every two years since 1986. A new sample of adolescent cohorts is scheduled to be drawn and interviewed in 1996.

IV. Recommendations

Our discussion thus far points to the need for routinely reported indicators of deprivation and dependence that describe both short and longer-run dynamic aspects of poverty and anti-poverty programs. In particular, our conceptual discussion points to the need for time series of indicators of:

- the number and characteristics of the "point-in-time" population of poor or dependent children or families with children;
- the number and characteristics of children experiencing first and subsequent transitions into and out of poverty or dependence and the events associated with these transitions;
- the number and characteristics of "long-term" poor or dependent children;
- intergenerational correlations of poverty and welfare receipt.

Our focus on children's indicators in this paper dictates that compilations of data on these kinds of indicators use children or families with children as the units of analysis. A family- or household-based analysis is problematic for longitudinal statistics since -- given composition changes such as a divorce, in which some children remain in the custody of one parent and other children are in the custody of the other parent -- it is ambiguous which new family is the "same" family as the old one (Hill and Duncan, 1985). Using individual children as the analysis unit solves this problem since they retain a unique identity across time. Furthermore, statistics on children can be compiled separately by developmental stage (e.g., 0-5, 6-11, 12-17), which is useful if research indicates a differing impact of deprivation or dependence by age.

Although fatally flawed by problems of data quality, CPS-based poverty indicators should serve as a model for how poverty indicators are processed and publicized. Compiled within six months of the completion of interviewing and reported at the same time each year, the CPS poverty indicators receive a great deal of publicity. It is crucial that all of the recommended indicators of short-run poverty or dependence be as timely and regular as the CPS poverty counts.

Furthermore, the methodology associated with the CPS poverty counts produces a reasonably consistent time series of poverty data. This is essential, given the difficulty in understanding and explaining the effects of changes in the methodology used in compiling the statistics. Finally, timely release of the CPS microdata files enables researchers to explore the robustness of the "official" indicators to various changes in definition and to produce tabulations of the poverty data that are better suited for particular policy concerns. The desirability of all social indicators should be evaluated with an eye toward these characteristics.

We now turn to detailed recommendations regarding these indicators. Our discussion assumes that the basic designs of the Current Population Survey and Survey of Income and Program Participation remain the same, with SIPP maintaining its intended 52-month duration.

A. *Short-run indicators of poverty*

1. Average monthly poverty rates and characteristics of the poor, published annually, based on data from SIPP. The month is the most appropriate accounting period for measurement of short-run poverty, although it makes sense to average the monthly poverty rates over a calendar year to smooth out seasonal fluctuations. At least some indicators should be provided using the child as the unit of analysis and separately by the child's developmental stage. Although rates of poverty are most important, indicators showing the degree of poverty (e.g., the average gap between the incomes of the poor and the poverty line) should also be compiled. These average monthly rates should replace the CPS annual poverty rates as the principal source of short-run poverty estimates. For purposes of historical comparisons, it would be useful to continue the basic CPS time series as well, although this should have a lower priority than furthering the timely release of the SIPP-based data.

2. Rates of transitions into and out of poverty, published annually, based on monthly data from SIPP, with at least some indicators using children as the unit of analysis. Methodological work is needed to determine the optimal measurement of an entry into or exit from a spell of poverty (e.g., does a single month out of poverty constitute a true "exit" from poverty?) Time-series data on the gross flows into and out of poverty will be invaluable in understanding the net changes in the average monthly rates. It should be noted that the current SIPP plans for nonoverlapping panels will make it impossible to construct a continuous time series of these indicators.

3. Events associated with transitions into and out of poverty, published annually, based on monthly data from SIPP, with at least some indicators using the child as the unit of analysis. These data should be coupled with the transition data listed above. It would be very useful to be able to track the marital, fertility and employment events associated with transitions into and out of poverty -- e.g., in the case of transitions into poverty: divorce/separation, the birth of a child to an unmarried woman, involuntary job loss, voluntary withdrawal from the labor force, cessation of transfer income payments. These events need not be defined to be mutually exclusive, since transitions may result from combinations of them. It should be noted that the current SIPP plans for nonoverlapping panels will make it impossible to construct a continuous time series of these indicators.

4. Income changes surrounding important demographic and employment events, published annually, based on monthly data from SIPP, with at least some indicators using the child as the unit of analysis. Past work has shown dramatic differences for ex-husbands, ex-wives and children in income changes surrounding divorce or separation (Duncan and Hoffman, 1985). These changes need to be tracked on a routine basis in order to monitor progress in child-support enforcement and other policies aimed at promoting an equitable burden following marital dissolution. Other candidate events include: *job loss*, with the attendant change in earned and family income and health insurance coverage, and *welfare-to-work transitions*, for which changes in total income and health insurance coverage are of greatest interest. The infrequency with which these events occur may require the pooling of several SIPP panels and less-than-annual reporting. It should be noted that current SIPP plans for nonoverlapping panels will make it impossible to construct a continuous time series of this indicator.

B. Long-run indicators of poverty

1. Distribution of poverty experiences over multi-year accounting-period "windows," published periodically, based on data from SIPP, with at least some indicators using the child as the unit of analysis. As argued by Gottschalk and Moffitt (1994) in the context of welfare receipt, a "total time in poverty" measure, when taken over a multi-year accounting period, provides a useful approximation of the distribution of short and longer-run experiences depicted in Figure 1. This should be taken once per SIPP panel, using as long a window as possible and compiled separately by developmental period. For a 52-month panel, this indicator would take the form of a distribution of the total number of months out of 52 that a child's household income was below the poverty line. Multi-year poverty and dependence indicators should be checked against and extended to longer accounting periods (e.g., the entire period of childhood in the PSID) using data from the PSID and NLSY.

C. Intergenerational indicators of poverty

1. Intergenerational poverty correlations. Both the PSID (e.g., Solon, 1992) and NLSY (e.g., Zimmerman, 1992) can be used to compare the parental economic status of adolescents with the economic status of those same individuals one to two decades later when the adolescents are well into their early-adult years. The design of the PSID now provides a substantial number of cohorts for whom intergenerational correlations can be calculated. Intergenerational correlations of poverty and earnings should be calculated and tracked periodically. An example of this would be the cross-classification of the years an individual spends poor during adolescence while living as a dependent against the years he or she spends poor as an adult.

D. Short-run indicators of dependence

1. Average monthly recipiency rates, published annually, based on data from SIPP, with at least some indicators using the child as the unit of analysis. As with poverty, the month is the most appropriate accounting period for measurement of short-run social-assistance receipt, although it makes sense to average the monthly recipiency rates over calendar years to smooth out seasonality. Data should be compiled separately by type of program (e.g., AFDC, Food Stamps, Supplemental Security Income) and for combinations of programs. U.S. Bureau of the Census (1992), Table 1, comes close to providing this kind of information.

2. Rates of transitions onto and off major social-assistance programs, published annually, based on monthly data from SIPP, with at least some indicators calculated using the child as the unit of analysis. As with poverty, methodological work is needed to determine the optimal measurement of the beginnings and endings of spells of social-assistance receipt. In contrast to the situation with poverty spells, there is some chance that the retrospective reports of social-assistance history can be used to classify transitions according to whether they are associated with first vs. subsequent spells of receipt.

3. Events associated with transitions onto and off major social-assistance programs, published annually, based on monthly data from SIPP, with at least some indicators calculated using the child as the unit of analysis. These data should be coupled with the transition data listed above. As with poverty, it would be very useful to be able to track the marital, fertility and employment events associated with transitions into and out of first and subsequent spells of social-assistance receipt.

4. Green Book-type indicators of point-in-time welfare receipt. Caseload records should be used to provide point-in-time indicators such as number of child recipients, demographic characteristics of recipient families, and fraction of recipients with no reported income other than AFDC. Virtually all of this information is available in SIPP, but for much smaller samples of recipients. Caseload data provide a valuable check on the reliability of the rates and trends estimated with SIPP data.

5. Take-up rates for major transfer programs affecting children, published annually, based on monthly data from SIPP, with at least some indicators calculated using the child as the unit of analysis. Since SIPP was designed to provide almost all of the information needed to determine program eligibility, it can be used to

monitor the fraction of children whose families qualify for various means-tested transfer programs but do not receive them.

E. *Long-run indicators of dependence*

1. Distribution of welfare-receipt experiences over a multi-year accounting-period "window", published periodically, based on data from SIPP, with at least some indicators using the child as the unit of analysis. As with poverty indicators, a "total time on welfare" measure, when taken over a multi-year accounting period, provides a useful approximation of the distribution of short and longer-run experiences depicted in Figure 1. This should be compiled once per SIPP panel, using as long a window as possible.

F. *Intergenerational indicators of dependence*

1. As with intergenerational poverty indicators, the PSID and NLSY samples should be used to calculate time series of associations of transfer-program receipt between parents and children. An example of such associations are presented in Duncan, Hill and Hoffman (1988), which tabulates, for a representative sample of females, the distribution of years between ages 14 and 16 in which parents received income from AFDC compared with the number of years between age 21 and 23 in which daughters themselves received AFDC.

G. *Experimental indicators*

1. Associations between family income and child outcomes. SIPP is experimenting with question modules focused on child development and may obtain periodic measurements of child health and cognitive development. It would be useful to track associations between household economic status measured between years one and four with children's outcomes measured at the end of the fourth year to see if income-outcome linkages were growing stronger or weaker over time.

The Fab Five

If forced to condense the above list to a handful of indicators, I would opt for the following:

- Short-run children's poverty: Average monthly poverty rates for children, published annually, based on SIPP (item A1, above).

- Longer-run children's poverty: Multi-year (e.g., 52-month) distribution of time in poverty for children, published as often as possible, based on SIPP (item B1, above).

- Short-run children's dependence: Average monthly AFDC recipiency rates for children, published annually, based on SIPP (item D1, above).

- Longer-run children's dependence: Multi-year (e.g., 52-month) distribution of time spent receiving AFDC income for children, published as often as possible, based on SIPP (item E1, above).

- Intergenerational dependence: Intergenerational correlations of welfare receipt, based on the PSID (item F1, above).

Bibliography

An, Chong-Bum, Haveman, Robert and Wolfe, Barbara. 1993. "Teen Our-of-Wedlock Births and Welfare Receipt: The Role of Childhood Events and Economic Circumstances." The Review of Economics and Statistics 75(2): 195-208.

Bane, Mary Jo; and Ellwood, David. 1983. "The Dynamics of Dependence and the Routes to Self-Sufficiency." Final report to the U.S. Department of Health and Human Services. Cambridge, MA: Harvard University, Kennedy School of Government.

Bane, M. J. and Ellwood D. T. 1986. "Slipping In and Out of Poverty: The Dynamics of Spells." Journal of Human Resources, 21:1-23.

Baydar, N., Brooks-Gunn, J., & Furstenberg, F. F., Jr. 1993. "Early Warning Signs of Functional Illiteracy: Predictors in Childhood and Adolescence," Child Development, 64(3):815-829.

Brooks-Gunn, J., Guo, G., & Furstenberg, F.F., Jr. (1993). "Who Drops Out of and Who Continues Beyond High School?: A 20-year Study of Black Youth," Journal of Research in Adolescence, 3(37):271-294.

Committee on Ways and Means. 1993. 1993 Green Book: Background Material and Data on Programs Within the Jurisdiction of the Committee on Ways and Means Washington, D.C.: U.S. Government Printing Office.

Corcoran, M., Gordon, R., Laren, D., & Solon, G. 1992. The Association Between Men's Economic Status and Their Family and Community Origins. Journal of Human Resources. 27(4):573-601.

DeParle, J. 1991. "To Moynihan, Welfare Dependency Signals New Ill" New York Times, December 9, 1991.

Duncan, G. 1984). Years of Poverty, Years of Plenty. Ann Arbor: Institute for Social Research.

Duncan, G., Brooks-Gunn, J. & Klebanov, P. 1994. "Economic Deprivation and Early-childhood Development. Child Development, 65(2):296-318.

Duncan, Greg J. Hill, Martha S. and Hoffman, Saul D. 1988. "Welfare Dependence Within and Across Generations." Science 239:467-471.

Duncan, G. & Hoffman, S. 1985. "A Reconsideration of the Economic Consequences of Marital Dissolution. Demography, 22:485-497.

Duncan, G.J. & Rodgers, W. 1991. "Has Children's Poverty Become More Persistent?" American Sociological Review, 56:538-550.

Duncan, G.J. & Yeung, J. forthcoming. "Extent and Consequences of Welfare Dependence," Child and Youth Services Review, 17: 157-182.

Ellwood, David. 1986. "Targeting 'Would-Be' Long Term Recipients of AFDC." Report prepared for the Department of Health and Human Services. Princeton, NJ: Mathematica Policy Research, Inc.

Furstenberg, F. F. Jr., Brooks-Gunn, J., & Morgan, S. P. (1987). Adolescent Mothers and Their Children in Later Life. Family Planning Perspectives, 19:142-151.

Gottschalk, Peter. 1992. "Is the Intergenerational Correlation in Welfare Participation across Generations Spurious?" Working Paper, Economics Department, Boston College.

Gottschalk, P. and Moffitt, R. 1994. "Welfare Dependence: Concepts, Measures, and Trends" American Economic Review, 84(2):38-42.

Haveman, R., Wolfe, B. & Spaulding, J. 1991. "Childhood Events and Circumstances Influencing High School Completion," Demography, 28(1): 133-157.

Hill, M. & Duncan, G. 1985. "Conceptions of Longitudinal Households: Fertile or Futile?" Journal of Economic and Social Measurement, 13(3-4):361-375.

Huston, A. 1991. Children in Poverty: Child Development and Public Policy, Cambridge: Cambridge University Press.

Lewis, Oscar. 1968. La Vida, a Puerto Rican Family in the Culture of Poverty: San Juan and New York London: Panther Books

Mathiowetz, N. 1994. "Autobiographical Memory and the Validity of Survey Data: Implications for the Design of the Panel Study of Income Dynamics," Survey Research Center, University of Michigan.

Mayer, S. 1994. "Measuring Income, Employment and the Support of Children," Harris School. University of Chicago.

Miller, J. and Korenman, S. 1994. "Poverty and Children's Nutritional Status in the United States". American Journal of Epidemiology. 140(3):233-243.

Solon, G. 1992. "Intergenerational Income Mobility in the United States," American Economic Review, 82(3):393-408.

Solon, Gary, Corcoran, Mary, Gordon, Roger, and Laren, Deborah. 1988. "Sibling and Intergenerational Correlations in Welfare Program Participation." Journal of Human Resources, 23(3):388-396.

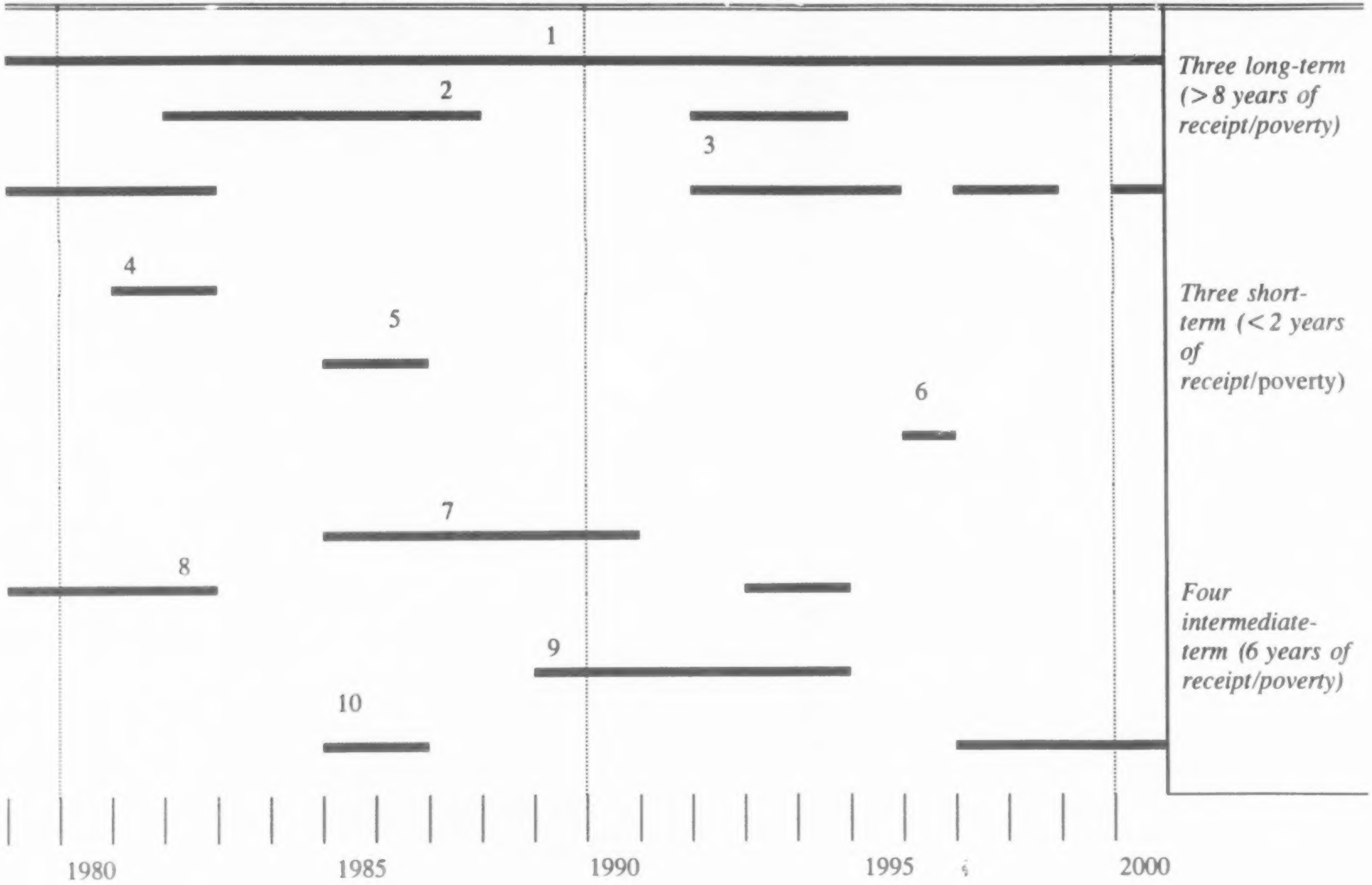
U.S. Bureau of the Census 1991. "Transitions in Income and Poverty Status: 1987-1988" Current Population Reports, P-70-24, Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census 1992. "Characteristics of Recipients and the Dynamics of Program Participation: 1987-1988" Current Population Reports, P-70-31, Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census 1993. "Poverty in the United States, 1992" Current Population Reports, P-60-185, Washington, DC: U.S. Government Printing Office.

Zimmerman, D. 1992. "Regression Toward Mediocrity in Economic Stature," American Economic Review, 82(3):409-429.

Table 1: Generic Longitudinal Patterns of Economic Deprivation -- e.g., Welfare Caseloads, Poverty, Unemployment



Endnotes

[1] For example, U.S. Bureau of the Census (1991, Table D-3) reports an overall poverty rate in 1988 (13.0%) that is 30% higher using data from the Current Population Survey than that calculated from the Survey of Income and Program Participation (10.0%).

[2] Using annual data from the Panel Study of Income Dynamics (PSID), Ellwood (1986) estimates that total years of lifetime AFDC receipt are distributed as follows: 30% last no more than 2 years, 40% last between 3 and 7 years, and 30% last 8 or more years. Estimates of poverty experiences are based on single spells rather than total years. Bane and Ellwood (1986) find that 60% of poverty spells last two years or less, 26% last between 3 and 7 years, and 14% last more than 8 years. Ellwood's (1986) estimates of single spells of AFDC show fewer shorter spells of AFDC than spells of poverty: 48% of AFDC spells last two years or less, 35% last between 3 and 7 years, and 17% last more than 8 years.

[3] Using annual data from the PSID, Ellwood (1986) found that 40% of first spells of AFDC were followed by second spells. Stevens (1994) finds that more than half of poverty spells were followed by subsequent spells within five years.

[4] For example, single-spell duration estimates of AFDC receipt based on monthly data from the Survey of Income and Program Participation show that many spells end within a single calendar year, e.g., 55% of spells of AFDC or other cash assistance programs for female-headed families ended within 12 months (U.S. Bureau of the Census, 1992, Table A-4).

[5] For example, Duncan, Hill and Hoffman (1988) show that the majority (66%) of daughters from highly dependent parental families did not, when in their early twenties, share the fate of their parents. At the same time, however, the fraction of daughters from highly dependent homes who themselves become highly dependent (20%) was much greater than the fraction of daughters from nonrecipient families who become highly dependent (only 3%). Mary Corcoran reports similar patterns using unpublished PSID data on intergenerational poverty.

[6] Mathiowetz (1994) provides evidence from a validation study that earnings cannot be recalled reliably for more than one calendar year, especially if earnings change substantially.

Parental Employment And Children

Judith R. Smith
Teachers College, Columbia University

Jeanne Brooks-Gunn
Teachers College, Columbia University

Aurora P. Jackson
Columbia University School of Social Work

Paper presented at the Conference on Indicators of Children's Well-Being, Washington, DC. November 18-19, 1994.

We want to thank the NICHD Network of Child and Family Well-being for their support in the writing of this paper.

Parental Employment And Children

The Changing Work Status Of Parents

The large numbers of mothers in the labor force are redefining family life and creating new social needs. Family role definitions, family division of time, child care arrangements and a woman's experience in both the work place and at home with her children may all be affected. Over 70% of all women with children ages 6 to 17 are employed outside the home and over half of women with children under one year-old are working (US Census, *Current Population Reports*, 1991). Although the employment of women with children outside of the home necessitates reliance on non-parental child care, federal legislation and marketplace solutions do not adequately address this reality, leaving individual families to balance work and family without adequate support. While there have been some recent incremental changes in policies which affect working families, such as the Family Leave Bill of 1992, the 1990 Child Care and Development Block Grant and significant increases in the Earned Income Tax Credit, a working mother in the United States is not yet assured of a paid parental leave, affordable quality child care or a choice to work shorter hours to accommodate child rearing responsibilities (Hyde, 1991; Kamerman, 1980, 1988; Kamerman & Kahn, 1981, 1991; Savarsky & Allen, 1984).

Social science research has also been affected by the absence of a coherent social policy response to maternal employment. Value conflicts about women's changing roles contributed to framing early research on maternal employment within a "social problem" matrix (Rallings & Nye, 1979). At times this research has been transformed into an ideological debate about whether or not maternal employment is "good" or "bad" for young children (Clarke-Stewart, 1988). Very little research has focused on the vicissitudes of how a mother's actual work experience impacts on the family's well-being or on the woman herself. Instead of investigating the various ways in which a mother's employment experience positively and negatively affects her children, most researchers have instead focused only on possible negative effects. For very young children, the concern is with the stress of coping with a mother's absence. For older children, the major focus has been the effect of reductions in parental monitoring or supervision. Surprisingly little research has investigated the impact of the particular types of jobs held by working women with children and the effect of the job itself on the mother, child and family. Instead, studies have focused on the effects of the job the mother is not doing (full-time child care). Now that more than half of all mothers are employed outside of the home, we need to move away from merely investigating maternal employment in terms of the effects of the separation on the child to a broader investigation of how the various aspects of a working mother's employment situation affect her sense of self, her parenting abilities and her time allocation. As Bronfenbrenner and Crouter (1982) point out:

> Throughout almost a half century of research on the working mother, almost no attention has been paid to the nature of her job . . work itself has been treated as an empty set, bereft of any structure or content that might be significant for the mother's role as a parent. (p. 41)

We do know that the economic situation of children whose mothers are employed is dramatically improved. During the 1980's the significance of a woman's contribution to family income became critical in offsetting the declines in men's earnings. If not for the increased work effort of their mothers, families of children in the poorest income group (the bottom 20% of the income distribution) would have lost 7.2% of their income compared to the actual loss of 2.5% during the recessions of the 1980's. Longitudinal studies show the importance of a mother's earnings in providing income to raise her family's income above the poverty threshold (Bane & Ellwood, 1983; Danziger & Gottschalk, 1990).

Because a mother's income contribution, particularly among low-income families, can significantly improve her family's economic well-being, the old question of whether or not mothers ought to be employed outside the home is not particularly relevant. Instead, the issue is under what circumstances will the financial benefits of being a working mother also lead to improved emotional well-being for the mother, child and/or the family.

Lost in most debates about employment is the role of the father. What are the effects of paternal employment on family functioning and child well-being. We know very little. The effect of dual-career marriages on men's career aspirations, the marital relationship and parenting responsibilities have been examined (Gilbert, 1985; Parke, 1982; Pleck, 1984, 1986). A father's employment has never been seen as a possible risk factor for child well-being, except in terms of job loss. Current demographic changes include relatively high rates of unemployment, underemployment and job loss for some groups of fathers (i.e. young minority men, men with little education, men employed in low wage jobs, as well as men losing jobs due to corporate downsizing) and the dramatic increase in divorce and never married families. Consequently, we believe that the effects of a father's employment (or lack of employment and lack of contribution to family income) on child functioning and adolescents' perceptions of the world of work must be further studied.

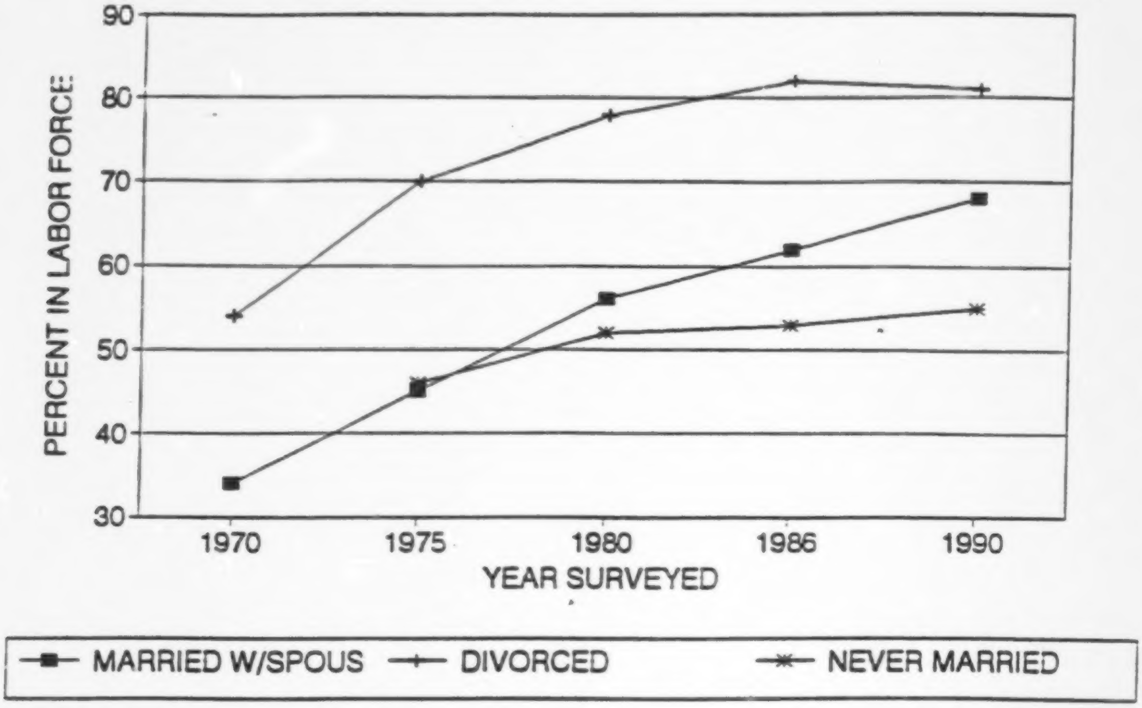
In this paper, we first review briefly the historical changes in mothers' work force participation, keeping in mind the importance of age of the child and marital status of the mother in interpreting rates of labor force participation. Then, the effects of maternal and paternal employment on the child are considered. In the next section, the three contexts most centrally influenced by working are identified. They are the parent's job, the home, and the child care environment. The following section reviews three national data sources vis-à-vis their inclusion of the measures identified for each of the three contexts. The paper concludes with a brief discussion of the relative importance of current changes in parents' employment on child well-being and a recommendation of the critical measures of parental employment for inclusion in future data collection efforts.

Women, Work And Mothering

Women's work outside of the home has historically challenged traditional ideas on sex roles and child care. With the development of a market economy and the move of production to outside of the family unit, women's status declined. The ideology of "true womanhood" and the science of "educated motherhood" emerged to define a woman's sphere within the home as "keeper of the hearth" and "heart" (Bernard, 1981; Cott, 1977; Ehrenreich & English, 1978; de Mause, 1974; Lopata, Miller, & Barnewolt, 1984; Oakley, 1974). In the twentieth century, findings from psychological and psychoanalytic research have been used by some to justify a woman's role outside of the market place or as a part-time or poorly compensated employee. The generally accepted belief within child psychology regarding the young child's need for a consistent relationship with one or two adults (Bronfenbrenner, 1979) has been used as a rationale to support the child's need for an at-home mother (Bowlby, 1951, 1969; Fraiberg, 1977). Feminist psychologists have pointed to the problems inherent in the gender-based assumptions regarding parenting and child care in our modern society (Chodorow, 1978; Dinnerstein, 1976; Scarr, Phillips & McCartney, 1990; Silverstein, 1991).

Despite the ideological support for an at-home unemployed mother, women's labor force participation rate has grown almost continuously since the Industrial Revolution. Even in the pre-industrial nineteenth century, women were typically not available for full-time child rearing as they had to work many hours at home doing farm and domestic chores. Women's labor force participation rates did recede temporarily after World War II when women were pulled out of jobs they held during the war to create jobs for the returning soldiers (Bergmann, 1986). The post-World War II period of economic growth and the baby boom led to a divergence from the trend of increasing women's employment. From 1950-1970, the modal family-type was a two-parent family with an employed father, and an unemployed mother/housewife available for full-time child care. Since 1970, this family type has become the exception. Figure 1 shows the change in the typical family, from the "traditional" family with only one earner (the husband) to dual-earner families.

Figure 1 **Trends in Numbers of Dual and Single Wage Earner Families: The Changing Labor Force Patterns of Families, 1940-88**



Source: *Monthly Labor Review*, March 1990, p. 18.

Mothers Who Work — The Demographics

The rates of increased labor force participation vary among demographic groups. Poverty and the lack of support from a father or husband have consistently motivated non-white, immigrant and low-income women to move into the labor force at higher rates than other women (Kessler-Harris, 1982). Yet, the differential between white and non-white women's labor force participation which was greater than 2:1 in 1900, is now almost even, with white women employed at slightly higher rates than non-white women (US Census, *Current Population Reports*, 1991).

Recessions and growing rates of unemployment and underemployment were partly responsible for the increased labor force participation rates of working married mothers since the mid-1970's. Yet, Barbara Bergmann (1986) demonstrates how explanations for women's entry into the labor force as based on "need" for money during an inflationary period is, in part, a way to rationalize women's employment in the public sphere and make it look as if the woman's choice is involuntary. If employment is involuntary, women are less likely to be blamed for abandoning their at-home role of child care provider. To classify women as merely "needing to" work masks the reality of the new economic relations and the accompanying changed gender roles of the twentieth century.

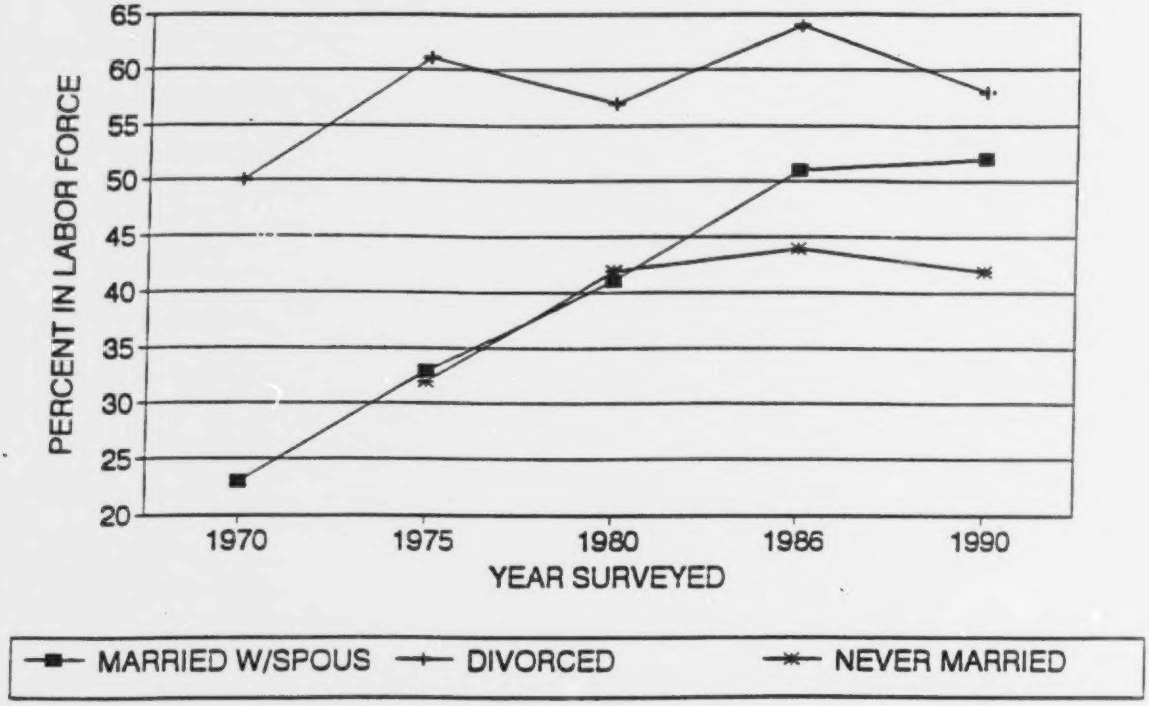
Factors that have led to women's steadily increasing labor force participation rates are highly influenced by structural economic and technological changes in the society, and accompanying changes in mores and expectations. Some of these changes have been: technological change and families desire for new consumer products, introduction of labor-saving devices in the home, growth of "suitable" occupations for women (clerical and service jobs), women's remaining at their jobs after marriage and childbirth and increasing wage rates which make staying out of the labor force a costly choice (Bergmann, 1986; Oppenheimer, 1979).

Employment Patterns Of Married Mothers

Despite the universality of women's increasing labor force participation across income groups, women married to husbands with lower earnings do have higher labor force participation rates than women married to husbands with higher earnings. The salary contribution of a low-income mother can determine whether or not her family lives in or out of poverty (Danziger & Gottschalk, 1990). In 1987, the poverty rate for white children under six-years-old in two-parent families was 2% if both parents worked compared to 18% if only the father worked. For black two-parent families, the poverty rate was 2% if both worked and 62% if only the father worked (US Census, *Current Population Reports*, 1991).

Although the age of a mother's youngest child affects mothers' participation rates, this has become less and less a constraining factor for married women. In fact, starting in the 1970's the greatest increases in labor force participation rates have been among married women with children under age one (see Figure 2).

Figure 2 Labor Force Participation Rates of Women with Youngest Children 0-3

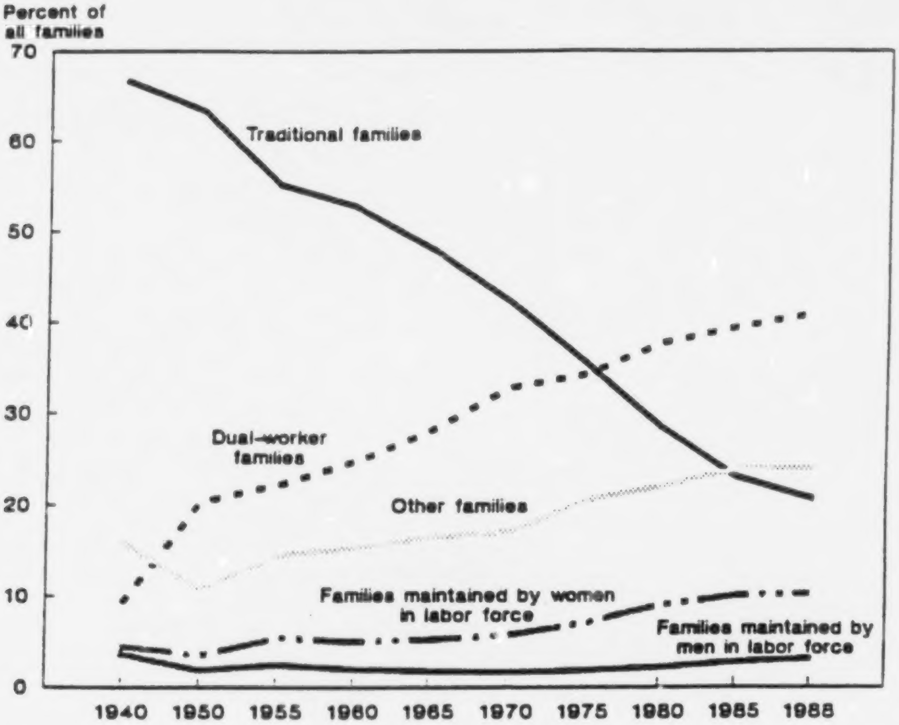


Source: Bureau of Labor Statistics, Bureau of the Census, *Current Population Survey*.

Employment Of Single Mothers

Divorced women with children have had the highest labor force participation rates of mothers with children. Figure 3 shows that beginning in the mid-1980's the labor force participation rates of married women with children started to approach that of divorced women. Never-married mothers have the lowest labor force participation rates. Never-married mothers tend to be on average very young, unskilled women with no previous labor force participation history. The lack of skills limits their eligibility to minimum wage jobs, which are inadequate to support their families; many therefore rely on AFDC. Demographic factors such as increasing numbers of single-parent families, the lack of sufficient child support and the low labor force participation of never-married single mothers have become a critical social policy problem because of the related poverty rates of these families. In 1990, almost two- thirds of the children under three years of age living in single-parent families were poor, compared to 12% of those in two-parent families (US Census, *Current Population Reports*, 1990)

Figure 3 Labor Force Participation Rates of Women with Youngest Children 0-18



Source: Bureau of Labor Statistics, Bureau of the Census, *Current Population Survey*.

Female-headed, single-parent families remain at a distinct disadvantage, even when the mother works full-time. Not only are these families dependent on only the one income, but the jobs available to most women offer salaries significantly lower than those available to men. Many women are not able to find jobs that pay sufficiently high wages to pay for child care and cover basic household expenses for themselves and their children. For those who are employed, working full-time year-round for single women heading families with children under six, decreased their poverty rates from 91.1% to 22.9% (US Census, *Current Population Reports*, 1991).

Effects of Maternal Employment on the Child

The question posed by developmental researchers has been whether or not children are influenced by a mother's employment outside of the home. Most studies have begun by examining the direct effects of maternal employment on children. A brief discussion follows. Alternative approaches include the examination of indirect effects of maternal employment upon children (as mediated or moderated by aspects of the child's or mother's life that are influenced by maternal employment) and family and community resource models, that focus on the resources available to the child.

Direct Effects

Studies have focused on effects on children in several age groups - infancy, preschool, and to a lesser degree, middle-childhood and adolescence. Outcomes of interest include verbal performance, school achievement, behavior problems, social relationships (especially with the mother and peers). In adolescence, some researchers have investigated outcomes related to aspiration and sex role identification. Various dimensions of the mother's work, beyond her absence from the home, are too infrequently studied -- timing, intensity, work preference, continuity, and strain between work and parenting roles. Little attention has been paid to income and its link with other work-related dimensions.

Generally, maternal employment per se is not consistently associated with negative outcomes for school-aged children (Gottfried & Gottfried, 1994). Indeed, maternal work seems to have benefits for adolescents (Hoffman, 1979). The only overall negative effects have been reported for young children. At least four lines of research focus on the effects of maternal employment and the young child. The first research tradition involves relatively small samples which have shown that children who experienced out-of-home care had a strained mother-child relationship as measured by the Ainsworth Strange Situation Paradigm[1] and were less well adjusted -- more aggressive with peers and less compliant to adult demands in their school-aged years (Ainsworth, 1964; Barton & Schwarz, 1981; Belsky, 1988; Erickson, Farber & Egeland, 1982; Haskins, 1985; Schwartz, Strickland & Krolick, 1974; Schwarz, 1983; Vaughn, Gove & Egeland, 1980).

Clarke-Stewart (1989) did a meta-analysis of the 17 studies which used the Strange Situation Paradigm to study the effect of infants of being placed in day care (at 12-24 months) or whose mothers were employed full or part-time. Her analysis showed that infants whose mothers are employed full time, compared with infants whose mothers do not work or who work part time, were more likely to be classified as insecurely attached. Yet, the percentage difference between children of employed and not employed mothers was relatively small, only 7%. Other researchers have concluded that these small effects are overall insignificant (Silverstein, 1991). In addition, Clarke-Stewart stresses that even if a mother's employment has an effect on a child's attachment rating, the real question is: What does this difference mean? She and others question the validity of generalizing the results of the Strange Situation Paradigm, an artificial laboratory experiment, to an assessment of the everyday mother/child relationship (Belsky & Steinberg, 1978; Vaughn, Dean & Waters, 1985).

In addition to questions raised about the validity of the Strange Situation Paradigm, many of the smaller studies on maternal employment and effects on children are limited methodologically in several other ways: (i) employment is measured via a gross variable of "employed" or "not employed" - without any indication of the number of hours the mother worked; (ii) samples are only generalizable to middle-class, two-parent, white households with school aged children; (iii) the mother's attitudes about her work are often overlooked and (iv) these studies are cross-sectional rather than longitudinal in design.

A second line of research, which corrects some of the above limitations listed above, is secondary analysis research done from national and larger scale studies, primarily the mother-child data of the National Longitudinal Survey of Youth (NLSY). The NLSY data set is longitudinal with detailed information on the mother's work experience over each year in the child's life (Baker & Mott, 1989; Chase-Lansdale, Mott, Brooks-Gunn & Phillips, 1991). One limitation of the data is that it includes no direct observation of the mother/child interaction. Instead there are child performance outcome variables such as the child's verbal ability (Peabody Picture Vocabulary Test -Revised PPVT-R, Dunn & Dunn, 1981) or maternal reports of behavior problems. All researchers using the PPVT-R as the dependent variable, except for Desai, Michael and Chase-Lansdale (1989), have found that a mother's hours of employment in the first year of the child's life have a unique and negative effect, albeit relatively minor effect, on the child's verbal facility (Baydar & Brooks-Gunn, 1991; Blau & Grossberg, 1990; Parcel & Menaghan, 1994; Smith, 1994). In addition, full time employment was found to have a stronger negative effect than part-time employment in the first year, as well (Baydar & Brooks-Gunn, 1991; Smith, 1994). However, no negative effects were found on the child outcomes for a mother's employment hours during the second or third years in the child's life. Blau and Grossberg found, in fact, that a mother's employment in year 2 raises the child's PPVT-R score somewhat. With behavior problems as the dependent variable, Baydar and Brooks-Gunn also found that mother's employment during the first year of life only had negative effects on a child's level of behavior problems (as reported by the mother).[2]

---

[1] Strange Situation Paradigm is a standardized laboratory experiment to assess the quality of the mother-child attachment through a procedure that studies the child's reaction to the mother's brief separation from the child. The experiment examines the child's interaction with an observer during mother's absence and the child's treatment of the mother on her return (Ainsworth, 1964).

A third approach has been to look at the child care arrangements of working mothers. Baydar and Brooks-Gunn (1991) and Smith (1994) using the NLSY mother-child data both found that type of care is an important explanatory variable in the maternal employment environment. Effects of child-care arrangements on the child's verbal facility (PPVT-R) vary with the gender and poverty status of the child. Informal care provided by a father or stepfather was found to be detrimental for the child's PPVT-R score (Smith, 1994) with stronger effects found for boys (Baydar & Brooks-Gunn, 1991) and poor families (Baydar-Brooks-Gunn, 1991; Smith, 1994). Child-care type also can lower the child's behavior problems, as rated by the mother. Informal care by a baby-sitter, grandmother or the mother herself resulted in significantly lower scores than care by the father or stepfather (Baydar-Brooks-Gunn, 1991).

The fourth approach of researchers has been to study how mother's employment changes family processes and gender relationships (Feree, 1990; Hoffman, 1974; 1989; Moore & Hofferth, 1979; Mortimer & Sorensen, 1984). For girls and boys a working mother who contributes to the family's income provides a role model that is different than that of a non- employed homemaker. A mother's employment may also change family relationships, with the father or the children assuming more responsibility for household chores and child care (Baruch & Barnett, 1987; Darling-Fisher & Tiedje, 1990; Gilbert, 1985; Gottfried & Gottfried, 1988, 1994; Hoffman, 1989; Manke, Seery, McHale, 1994; McHale & Crouter, 1992; Pleck, Staines & Lang, 1980).

Indirect or Mediating Effects

While few direct effects of a mother's employment have been found on child well-being, indirect effects via the mother's emotional well-being have been described. Theoretically, combining work with family roles has been associated with detrimental effects and beneficial effects. For example, Goode (1960) argued that the more roles people accumulate, the more likely they are to encounter incompatible expectations or excessive demands on their time and energy (role strain). The further assumption of this perspective is that role strain erodes psychological well-being. Others (Barnett & Marshall, 1992; Baruch & Barnett, 1987; Marks, 1977; Sieber, 1974) have challenged this view, arguing instead that multiple roles can enhance well-being by offering multiple opportunities for increased status, privileges, and self-esteem, particularly when people are committed to the roles they occupy. Studies concerned with the relationship between women's psychological well-being and paid employment have looked at the mediating variables of social support, marital status, type of job, number of children in the household and preference for employment (Cleary & Mechanic, 1983; Jackson 1992, 1993, 1994; Kessler & McRae, 1982; Radloff, 1975). Positive effects of employment on women worker's mental health have been found. Work provides social contact, a sense of identify and a feeling that one is needed by others (Jahoda, 1982). Employment has been found to lead to negative consequences for some women who experience overload either due to heavy demands within the job or due to multiple role strain. Multiple role strain can occur for women who have difficulty locating affordable quality child care, or a spouse who disapproves of a wife's employment, or a young child under one year old (Baruch & Barnett, 1987; Hochschild, 1989; Hoffman, 1979; Piotrkowski & Katz, 1982; Moen & Dempster-McClain, 1987; Pleck, 1984; Repetti, 1987; Staines, 1980; Ross, Mirowsky & Huber, 1983; Walker & Best, 1991).

---

[2] Belsky and Eggebeen (1991) using a created variable of emotional maladjustment from mother's ratings of the child's temperment found no negative effects in year one, but significant negative effects from mother's full time employment if begun sometime during the first two years of life. Smith (1994) using Belsky and Eggebeen's maladjustment variable on a younger sample of children did find negative effects for year one.

A mother's level of satisfaction with her role has been the most extensively studied mediator between maternal employment and child development. Many studies have confirmed that a mother's satisfaction with her role, whether she is employed or not, has positive effects on her children (Baruch & Barnett, 1986; Farel, 1980; Gove & Zeiss, 1987; Guidubaldi & Nastasi, 1987; Hock, 1980; Hoffman, 1989; Ross, Mirowkshy & Huber, 1983; Spitze, 1988). In contrast, dissatisfaction with the maternal role is associated with negative effects on children — both in the school adjustment (Farel, 1980; Woods, 1972) and behavior problems (Barling, Fullager & Marchl-Dingle, 1986; Forehand, McCombs & Brody, 1987; Jouriles, Murphy & O'Leary, 1980; Hock, 1980; Lerner & Galambos, 1985). The underlying hypothesis is that a woman's feelings of self-fulfillment influence her functioning as a mother and affect what is mediated to her child through her child-rearing practices. Improved maternal self-esteem is hypothesized to lead to positive mood changes in the mother, more acceptance by the mother of her child, and more sensitive mothering.

Reviews of this evidence have concluded that although employed women in comparison with their non-employed counterparts are in somewhat better mental health, maternal preferences with regard to employment are a significant factor in this relationship (Gove & Peterson, 1980; Kamerman & Hayes, 1982; Spitze, 1988). However, since this research is based largely on samples of middle-class, married, white women, little is known about the effects of early employment on the psychological well-being of poor, single, and minority mothers, or the processes linking their employment to developmental outcomes for their children. Jackson's (1992, 1993, 1994) recent studies of psychological well-being in a sample of single, employed, black mothers of preschoolers found that mothers who preferred their current employment status, while no less depressed, were lower in role strain than their employed counterparts who preferred to stay home. Marshall and Barnett (1991) found that there were clear gains for all women who combine work and family, but that the strains were particularly intense for working class women because of their limited resources to help them with the double shift of combining work and family. McLoyd's (1994) study shows the positive effect of a mother's perception of instrumental help from others in a sample of poor African-American single parent families, as well as the negative effect of current unemployment on the mother's emotional well-being and her parenting style.

The type of job the mother performs has also been investigated as mediating between the mother's employment and her functioning at home with her children. Jobs which encourage autonomy and self-direction have been shown to affect the mother's intellectual flexibility and positively affect the mother-child interaction at home (Miller, Schooler, Kohn & Miller, 1979; Menaghan & Parcel, 1990; Parcel & Menaghan, 1994). The lack of opportunity for less educated women to acquire jobs which encourage autonomy and self-direction may therefore have an effect on their children.

Family and Community Resources: A Framework for Studying Effects of Employment

Different frameworks have been proposed to explicate the links between the various contexts in which children reside and children's well-being. Developmentalists have favored models focusing on the interplay of various ecosystems (or contexts), on the contribution of risk and protective factors, and on the socialization practices of the family, school, and peer group (Bronfenbrenner, 1979; Garmezy & Rutter, 1983; Maccoby & Martin, 1983; Bornstein, in press). More economically-oriented frameworks have focused on resources which are available to children (Haveman & Wolfe, 1991, 1994). And more sociologically-focused frameworks often add social capital and networks to their equations (Coleman, 1988). All of these frameworks have been used (to varying extents) in the investigation of the effects of parental employment, job loss, and unemployment upon children and adolescents. However, the data collected (and the type of sample used) vary somewhat across disciplines. Thus, we know a great deal about mother-child relationships around a year of age vis-à-vis maternal employment; however, information is based on a series of small scale studies (40 to 100 children) of primarily white middle class families (Belsky & Steinberg, 1978 ; Clarke-Stewart, 1989). Likewise, data on the links between income, single parenthood, and maternal employment are quite extensive for outcomes such as school achievement (Krein & Beller, 1988; McLanahan, 1985; McLanahan & Sandefur, 1994). However, little is known about the associations between maternal employment and family processes as they influence children, or about how these associations might be affected by the age, health, or gender of the child, the age, education, or marital status of the mother and father, the type and intensity of work performed.

We believe that more micro-analytic and macro-analytic perspectives need to be integrated if we are to go beyond a mere description of parental employment patterns. To this end, a family and community resource framework is adapted here (see Brooks-Gunn, Brown, Duncan & Moore, 1994 for a more complete explication of this model, which is based on the work of Haveman & Wolfe, 1994, and Coleman, 1988). At least four categories of family resources are identified--income, time, human capital, and psychological capital resources. The last category includes many of the so-called "process" variables--parenting behavior, parental attitudes and beliefs, parental emotional health, social support. Parental employment potentially may influence all four family resources, which in turn affect child outcomes. Thus, we expect that parental employment will have many more indirect than direct effects upon children. Understanding how families allocate resources within the family is critical to the specification of policies related to enhancing the well-being of children whose parents work.

Community resources include institutions such as schools and child-care settings, as well as income, human capital, and social capital resources. Most studied in terms of maternal employment is the availability of schools and child-care. Also important are the social networks that help parents to find adequate child care, to lobby for more or better child-care, and to locate jobs (or jobs with adequate benefit packages).

The intersection between family and community resources also needs to be specified (Brooks-Gunn, in press). This is especially true when considering maternal employment which is dependent on child care. Additionally, little work has considered the characteristics of the child that may influence the ways in which maternal employment and family or community resources interact. For example, some but not all work suggests that infant boys are more likely to be affected by maternal employment than girls (Belsky, 1988; Chase-Lansdale & Owen, 1987: Desai, Chase-Lansdale & Michael, 1989).

Family resources will be briefly discussed here vis-à-vis what is known about their links with employment and child outcome and their potential value as mediators or moderators of the association between parental employment and child outcome. The **income** brought into the family (or economic setbacks from job loss or unemployment) has the most obvious and potent effect on the quality of family life and child well-being. Research suggests that family income is one of the most important factors in the young child's environment and is related to the adequacy of prenatal care (Kalmuss & Fenelly, 1990), low birth weight and infant mortality (Klerman, 1991), cognitive and socio-emotional development (Duncan, Brooks-Gunn & Klebanov, 1994; McLoyd, 1990), physical health (Miller & Korenman, 1994), level of school readiness (Copple et al., 1993) and rates of adolescent high-risk behaviors (Dryfoos, 1991). Inadequate income, unemployment or economic strain can affect children indirectly by influencing the parent's well-being which then affects their attitudes and the quality of the parent-child interaction (Crinic, 1983; Conger, Yang, Lahey & Krupp, 1984; Elder, LIker & Cross, 1984; Elder, Nguyen & Caspi, 1985; Kohn, 1969; McLoyd 1990, 1994; Pascoe & Earp, 1983; Tulkin & Kagan, 1972). The stress associated with low-income may severely limit the emotional energy mothers have to invest in parenting. Economic hardship has been shown to place women with children at high risk for depression (Belle 1982, 1990; Hall, Williams & Greenberg, 1985; Pearnlin & Johnson, 1977). A mother's depressive symptomatology is a frequently used indicator of a mother's mental health. Maternal psychological distress, in turn, has been shown to be a significant mediator between economic hardship and child developmental outcomes through its effect on parenting (McLoyd, 1990, 1994).

Time is a second important parental resource which is affected by a parent's employment situation. If both parents are employed outside of the home, a critical question is how does this affect the quality and quantity of the time that the child gets to spend with the parents and other caregivers. Family activities can include sharing play and leisure time, eating together, doing housework, and educational activities like reading or watching a movie. Time-use diaries have been successfully used to describe how much time is spent by parents in child-oriented activities (Timmer, Eccles & O'Brien, 1985). Yet, there is no on-going nationally representative data base that includes time diaries of families with children. Time spent with other caregivers can be educational and/or nurturing, or lacking in adequate stimulation or developmentally appropriate caregiving. A large body of research has documented the importance of quality of child care type on child well-being.

A **secure emotional base** for the developing toddler is thought to be a critical psychological resource. Developmental frameworks document the existence of stages in the young child's development as it occurs within the mother/child relationship (Bowlby, 1969, 1973; Freud, 1905; Erikson, 1959; Piaget, 1937; A. Freud, 1944; Kohut, 1971; Kernberg, 1967; Greenspan, 1981; Mahler, Pine & Bergman, 1975; Stern, 1985). All of these schemas describe the emergence of a focused attachment to the mother (or primary caregiver) beginning during the second six months of the child's life which follows a developmental timetable. Developmental milestones in the child have been shown to emerge within, and be influenced by, the maternal relationship. The documentation of phase-specific developmental achievements by the very young child within the relationship with the parent raise the question: Will these achievements be delayed or impaired if mother is absent for some part of the day?

DOMAINS OF STUDYING MATERNAL EMPLOYMENT

To understand both the intersection of work and family and the possible effects on the family system, the multiple contexts in which parents and children operate must be identified (Bronfenbrenner, 1986). The three most important are the work environment, the home environment, and the child care environment. Each context will be discussed separately. Measures of these three environments will be identified. In a later section, three data sets will be reviewed with an eye to their inclusion of each domain and specific measures within each domain.

Parental Employment Conditions on the Job

Understanding the effect of a parent's employment situation needs documentation of data on several aspects of the parent's job. We see the need for data on employment hours of both parents. Child well-being will be affected by the inputs of both parent's employment, as will child care arrangements and time allocation in the family. We will focus first on the data needed on the mother's employment situation and then point out how the data needs would be similar or different for describing the father's situation. It is assumed that these domains will have direct and indirect effects on child well-being. Direct effects can be measured in terms of the associated economic and emotional well-being of the family and the child. Indirect effects can be measured via the parent's emotional availability to the child and parenting behavior.

Mother's employment situation

Timing and stability of parental employment. The mother's age at first birth, her prior employment experience, her educational achievement and number and age of other children is associated with the likelihood and strength of her ability to contribute to the family's economic well-being. The employment experiences and educational level of women prior to childbirth are important indicators of the strengths of a woman's labor force attachment. Years of employment prior to childbirth and work status during pregnancy have been shown to be predictive of the timing or reentry into the labor force after childbirth as well as providing some data on mother's access to higher paying jobs, health insurance to cover prenatal care, childbirth and well-baby care and disability insurance coverage for a paid short-term maternity leave. Young women with no prior labor force attachment, and a low level of educational attainment, are less likely to seek employment after childbirth and may remain out of the labor force for many years and therefore their children will be at greater risk for poverty under these circumstances (Bumpass & Sweet, 1980; Leibowitz, 1974).

Information needs to be collected in order to study the effects of a mother's employment on : (i) length of time mother spent out of the labor force with her infant or toddler (if any) or age of child when mother began employment; (ii) intensity (number of hours worked each week) during each year of the child's life; (iii) marital status and number of other adults in the household; (iv) number of job changes the mother made over the year (stability); and (v) proportion of family income provided by mother's income.

Longitudinal data are important to capture the timing and intensity of employment, the age of the child when a mother returns to or begins employment, how employment hours change (if they do) as children enter school or preschool, and whether work schedules include seasonal fluctuations. Some researchers have found that some dual-earner families become single-earner families during the summer (Crouter & McHale, 1993). Gathering data on a workers' daily and yearly schedule may lead, however, to more variation than is interpretable. As many jobs include shift work, or seasonal fluctuations, a focus on the worker's feelings about their schedule, rather than detailed schedule information beyond total hours worked per week, may be most profitable (Presser, 1989).

Salary. A second and obviously critical aspect that needs to be measured of a mother's employment situation is her salary. Knowing a mother's hourly wage is necessary to understand how much her hours outside of the home are providing economic benefits to the family. How much income is necessary to compensate for the possible stress the child might experience from coping with a mother's absence? Smith (1994), controlling for many family and maternal characteristics, found that a mother's salary of $23,000 could offset the negative effects on the child's verbal facility (PPVT-R) from the mother's 40 hours of employment per week in the first year. This implies that most low-skilled mothers working at a minimum wage job would not be able to earn a high enough salary to offset the negative effects of the mother's absence and that their children would be at higher risk for negative effects from early employment. Racial differences in wages will also affect children's economic status and may be more detrimental than maternal employment per se (Eggebeen & Lichter, 1991).

Fringe Benefits. Data on access and availability to various packages of fringe benefits are critical for our understanding of how effective a job is in meeting family and individual needs. Employment benefits have been termed the "new property". Benefits accounted for 27% of employee compensation in 1989 compared to 17% in 1966. For those who work, pensions are a principal form of wealth, providing greater value for middle-aged Americans than the once prized family home or the automobile. Lack of benefits in the majority of jobs available to women who work in part-time employment increases the inequality between children in these families compared to those whose parents work at jobs which provide a full menu of benefits. With the increasing likelihood of marital disruption, a good job may become more central to economic security than family relationships (Drucker, 1976; Reich, 1964; Glendon, 1981; Kamerman & Kahn, 1988). Information on a workers' benefits should track whether the individual receives a pension, health insurance, paid vacation time, maternity or parenting leave, flexibility of work schedule, child care vouchers or services and counseling services.

Employment can offer not only the possibility of employer-provided benefits but also the government concomitants of employment - including social security, disability insurance, unemployment insurance and Medicare. Social security coverage is a critical benefit of employment, yet most survey questionnaires do not differentiate whether a worker is being paid on or off the books - (i.e. whether the employee is eligible for social security). For those women not in the labor force, it is important to know if they and their children are covered by the father's benefit package or by statuary benefits such as Medicaid or Medicare.

Occupational Complexity. Measures of the mother's working conditions in terms of the level of routinized or occupationally complex working conditions may be important. Some studies have shown that parents encourage the styles of behavior that are rewarded in their own line of work (Miller, Schooler, Kohn & Miller, 1979; Parcel & Menaghan, 1990, 1994; Schooler, 1987). Kohn's social structure and personality framework are applicable when studying the impact of mother's employment conditions on her care of her children when she is with them. Parental occupational complexity and opportunities for self-direction and autonomy on the job are the critical dimensions of parental working conditions that influence child-rearing values and behaviors (Kohn, 1969; Kohn & Schooler, 1973, 1982; Parcel & Menaghan, 1994). Parents in high-complexity occupations place less emphasis on direct parental control, instead their parenting style promotes the child's internalization of parental norms. When internalization is successful, children use these internal standards to monitor their own behavior, reducing the frequency of "acting out" behavior and the necessity for parents to impose external control. All data sets that measure a mother's job by Census bureau categories can be linked with the Dictionary of Occupational Titles to assess levels of occupational complexity (Parcel, 1989).

Job satisfaction. Job satisfaction is affected by economic compensation, the physical environment, as well as by relations with supervisors and fellow workers. Opportunities for social contact can be particularly beneficial for single mothers with few other sources for social support with possible cross-over effects to their children (Parry, 1986; Warr & Parry, 1982; Repetti, 1989). Rauh (1994) has shown the positive effect of job-linked social networks for pregnant women and mothers of young children. Working in physically dangerous or unhealthy conditions affects a workers' well-being. Salary, experience, and opportunities for promotion is a critical part of job satisfaction, particularly for women whose income is necessary for their families' well-being (Feldberg & Glenn, 1979; Loscocco, 1990; McKenry & Hamdorf, 1985; Martin & Hanson, 1985). More work is needed to understand what job satisfaction is really tapping. A survey done by the Women's Bureau of the Department of Labor of 250,000 working women found that while 80% of the women reported that they loved their jobs, nearly half of them reported feeling that they were underpaid because they were women and under stress because of managing work and family (Lewin, 1994).

Preference for employment. With changing expectations for women to hold dual roles - as workers outside of the home and as homemakers, data are necessary to track women's satisfaction with their dual or single role. We need to know about a women's level of satisfaction in terms of being a worker and also her satisfaction in terms of facilitating relationships with her children and family. It would be informative to know whether an employed mother would prefer to be working less hours or not working at all. Likewise, a mother's report of the emotional and concrete help she receives from a spouse or other household members is an important consideration in her response to the maternal employment situation.

Father's employment situation

The research on the effects of a father's employment situation on child well-being are very limited. Several aspects of the father's work have been shown to be important, however. One focuses on the effects of job loss on the family and on children. Elder et al (1985, 1986) looked at the effects of father's unemployment during the Great Depression; and Conger et al (1992) have looked at unemployment in mid-western farm communities during the 1980's. Both studies found that potential job loss and unemployment are associated with instability, hostility and inconsistent parenting on the part of father, and these behaviors are linked to less optimal child and adolescent outcomes. Additionally, fathers who were more unstable prior to job loss were most likely to show very high levels of negative parenting when a job crisis occurred (an accentuated effect); children in such families have the worst outcomes. Interestingly, maternal parenting behavior was not particularly predictive of child outcomes, and was less likely to be influenced by the job loss of the father. Whether or not similar links would be found in families where the mother is the primary wage earner needs further study(McLoyd, 1990, 1994).

Another line of research focuses on the complexity of the parent's occupation. Kohn's (1969) work found that fathers with jobs that encouraged self-direction, encouraged independence in their children rather than conformity (Kohn & Schooler, 1982; Schooler, 1987). The father's salary is of obvious importance in determining the family income and socio-economic status of the family and associated child well-being. Parcel and Menaghan (1994) are the only ones to date to investigate the effect of a father's work schedule (hours) on child outcomes. They found that for children under three, fathers' working less than full time was associated with elevated behavior problems. They suggest that "fathers' work schedules may be important pathways through which children absorb appropriate behavioral norms and develop verbal skills that serve as the foundation for future cognitive attainment" (p. 1003).

Home environment

There is a large body of research which connects children's home environments and their health and development (Bradley & Tedesco, 1982; Clarke-Stewart, 1973; Wachs & Gruen, 1982). Employment may affect the child's home environment in several ways. Increased income may allow the parents to buy additional educational toys or books which increase the cognitive stimulation in the home. Employment might also increase the parents' cognitive functioning which, in turn, will affect their interaction with the child. Working may also affect the parents' mood when they are at home, thereby increasing or decreasing their emotional availability for interaction with the child. Finally, employment might affect the parents' resources to create a safe, well-ordered and clean home environment.

Provision of Learning Experiences and Responsivity

A large number of researchers studying the relationship between the home environment and child development rely on the Home Observation Measure of the Environment (HOME) scale -- a standardized measure of the environment, which was originally developed to identify and describe homes of infants and children who were at significant developmental risk (Bradley, Caldwell, Rock, Hamrick & Harris, 1988; Elardo & Bradley, 1981). The full HOME scale taps cognitive variables including language stimulation, provision of a variety of learning experiences and materials, and encouragement of child achievement; social variables include the parents' responsiveness and warmth; and a measure of the physical dimensions of the home including cleanliness, safety and amount of sensory input. Several researchers have investigated how maternal employment and paternal employment influence the child's home environment using the HOME scale. Several constructs in the HOME scale have been linked to child outcomes: maternal warmth and responsivity, variety of learning experiences, safety of the physical environment, father involvement and the level of parental punitiveness (Brooks-Gunn, Klebanov, Liaw & Spiker, 1993; Gottfried & Gottfried, 1988). Menaghan and Parcel (1991) found that maternal working conditions influence the strength of a child's home environment. Those mothers who worked in occupations with more substantively complex work activities created home environments that were more cognitively enriched and more conducive to socio-emotional development.

Time allocation

Employment affects the time availability of the mother and father for family activities and limits leisure time. Mother's time in the labor force is often taken as a possible problematic indicator of time not available for parenting. Collection of data on the effect of dual-parent or single-parent working families on time allocation in family members' lives is much needed. While employment may bring in additional income to the family, it may also create "time poverty" -- a deficit of social time for shared family activities, leisure time or household chores. How children spend their time is an important indicator of their well-being (Task Force on Youth Development and Community Programs, 1992). Information is needed on: (i) the amount and nature of time parents and other caretakers spend with children; (ii) the amount of time older children and adolescents spend in unsupervised activities; and (iii) the amount and nature of household activities each member of the family is performing. The Michigan Time Use Studies successfully obtained detailed information on family members' time use (Juster & Stafford, 1985). These time diaries were able to provide information on family processes by documenting the division of labor within the family and demonstrating the role of a gender, marital status, educational attainment and employment status on time allocation for household chores, leisure time or television viewing. They found, for example that single mothers spend more time in employment than married mothers, that employed mothers spend much less time on housework than non employed (single-earner) mothers and that college-educated parents spend more time reading to their children than lesser educated parents (Timmer, Eccles & O'Brien, 1985). No other national data sets have such data. Maternal education is often used as a measure of the likely quality of the mother's time with the child. We see the collection of time diaries as an extremely useful, although costly, addition to a national data collection effort.

Smaller studies are doing innovative research tracking parental time allocation in terms of effects on monitoring of school-aged children's activities and allocation of household chores between spouses and among children (McHale, Bartko, Crouter & Perry-Jenkins. 1990; Manke, Seery and McHale, 1994). Manke et al. found that fathers in dual-earner families performed more housework than fathers in single-earner families and that girls did more housework than boys, with girls in some families substituting for their father's household tasks. A planned sibling design funded by NICHHD, Crouter and McHale will be able to track within family variation on time allocation and chore division focusing on possible gender and temperament differences of the children.

Changing gender relations

Changing labor force participation rates of women affects women's and men's role definitions. Women's greater economic independence could influence children's images of what men and women do and can become. Some researchers rather than only investigate the effect of a mother's salary, instead look at the gap between a mother's and a father's income in dual-earner families. The hypothesis is that the gap or lack of gap between spouses' income can be predictive of marital equality or inequality in terms of the power that the lesser earner has in family decision making (McHale & Crouter, 1992). How this might affect children's well-being is not known.

Child Care Environment

When a child's parents are employed, alternative child care arrangements must be provided for the young child. The quality of this care may have a direct effect on child well- being, as well as have a spillover effect to the parent's job satisfaction and "peace of mind". A critical question is whether the quality of the child care enhances or undermines the effects of parental employment on child well-being. This section will be brief, as child care is discussed in the paper by Deborah Phillips.

Longitudinal studies done in Sweden, where there is universal high-quality affordable child care, found positive effects for early entry to substitute group care for cognitive and social development. No equivalent comprehensive longitudinal data exist on effects on children of different types of child care in the United States.[3] Descriptions of child care arrangements do not provide a picture of quality of child care. Research has demonstrated repeatedly that quality of care varies tremendously between and within states, depending on licensing and regulatory controls (Morgan, 1987). The most commonly used type of care for pre-school children is unlicensed informal care by a relative or non-relative (family day care). This popular form of care, however, is the least studied in terms of effects on child well-being. Smith (1994) using the NLSY mother-child data found that there were strong negative effects on a child's verbal facility if the mother was employed and the child was cared for in informal relative care by a father, stepfather or sibling. Galinsky, Howes, Knotos & Shinn (1994) also found negative effects for child care on young children if the care was provided by a relative in the relative's home. Much more information is needed about the informal and formal child care environments of pre-school children, as well as the arrangements being made for school aged children after dismissal from school. When gathering data on the effects of the child care environment on child well being we suggest questions on: (i) person providing the care and place of care; (ii) ratio of children to adults; (iii) timing of entry; (iv) education and training of provider (v) stability of care or number of changes.

The federal data which is now available include the 1990 National Child Care Survey (NCCS), the Profile of Child Care Settings and data from the Current Population Survey. The NCCS provides a cross-sectional picture of the child care arrangements of children under age 13 in a nationally representative sample of families (Hofferth, Brayfield, Deich & Holcomb, 1991). The Profile of Child Care Settings presents information on the supply of child care provided in public and private child care centers, nursery schools and preschools, as well as regulated family day care homes (Kisker, Hofferth, Phillips & Farquhar, 1991). The Current Population Survey (Household Survey) obtains data every other year on whether children 3- and 4-years of age are enrolled in a nursery or day care center with some educational component. CPS data indicate that preschool participation was low for poor children. Only 35% of poor 3- and 4-year-olds attend preschool. Participation rates are particularly low for children in immigrant families, and children in rural households. In addition, only 29% of eligible 3- and 4-year-olds attend Head Start (General Accounting Office, 1994).

---

[3] The National Institute of Child Health and Human Development Study of Early Child Care is currently in the data collection stage. This study will make a significant contribution in providing detailed longitudinal data on the experience of children and their families in a variety of child care settings through the entry into kindergarten. The data will also include many key measurements of the parent's employment situation. This data, however, is not nationally representative and access to the data is restricted to the investigators.

DATA COLLECTION EFFORTS

In considering priorities for collection of social indicator data on parental employment, there are three levels of variables that are important to monitor. The primary level of needed data is related to tracking the employment trends of each of the child's parents. The second level includes monitoring related changes in the family and child care environment. The third level includes effected child outcomes. We have examined three large longitudinal data sets vis-à-vis the data collected related to these three levels. The data sets are the National Longitudinal Survey of Youth mother-child data (NLSY), the Panel Study of Income Dynamics (PSID), and the Current Population Survey (CPS). The NLSY mother-child data is a supplement to the annual NLSY survey begun in 1979. The original sample included over 12,000 young men and women ages 14-21. The sample includes a special military subsample, as well as oversampling for blacks, Latinos, and poor whites. The data includes rich detailed information on the respondents labor force participation for each week beginning in 1979. Parents are interviewed in each year since 1979 about hours, salary, fringe benefits and some job satisfaction measures.

In 1986, child assessment measures were given to the children of those women who had become mothers (n=2,918). Child assessment measures have been continued on a bi-annual basis. The sample, however, is not nationally representative, as it only represents those women who gave birth in the early phase of their employment careers. The 1992 child data will include assessments on the children of 70% of childbearing women (children of women in the youth cohort who delayed childbearing into their late thirties will be included when they give birth).

The PSID is a survey which has been conducted on an annual basis since 1968. In 1993, the survey involved some 7900 households, with an over-sampling of black and Latino families. A strength of the data set for studying the effects of parental employment is its extensive and detailed information on family material resources and the transfer or income on a monthly basis. A strong limitation of the data for studying the effects of parental employment on children is that there is only very limited child outcome measures on children before they turn 16 years old. Yet, one can trace the long term effects of a parent's employment situation on adolescents or young adults (16+).

The Current Population Survey is intended to provide estimates of employment, unemployment and general characteristics of the labor force. Monthly labor force data is collected for the nation, 11 of the largest states, New York City and Los Angeles. The total sample size is approximately 71,000 households. About 57,000 households are interviewed in the monthly survey. The data set was not intended to study the effects of employment on children, but demographic data has been collected on the children in the adult respondents' household beginning in 1979 and one can get detailed estimates of the types of jobs parents are holding, their hours, unemployment spells, fringe benefits etc.

Table 1 describes which domains of parental employment are currently being monitored in these three data sets. Table 2 describes available data on the family and child care environment. Table 3 describes available child outcomes. Table 4 includes data on background parental resources. Examination of these tables illustrates that (i) on the one hand there is sufficient data currently being collected on maternal employment to allow for immediate monitoring of mother's employment situation; and (ii) there are many gaps in our data collection efforts particularly related to father's employment and child outcomes. Child outcomes are only available in the NLSY, and this sample is not only nationally representative of younger mothers and their children. Detailed parental data are only available if the father lives in the household and is married to the mother. We can therefore know little about the effect of father's employment (or non-employment) on children from divorced or never married families. Other limitations include: job satisfaction measures for both mother and father are limited to a global measure collected for moms each year and data collected on occasion about other domains. None of the three data sets allows for measurement of occupational complexity, but they all have census codes which can be merged with the Dictionary of Occupational Titles which can measure occupational complexity. Measurement of the home environment, is primarily captured in the NLSY with the HOME scale. None of the data sets have time diaries. Child care arrangements are most fully described in the NLSY. Parental resources for the mother are fairly well documented in all of the three data sets, with one serious limitation. Mother's depressive symptomatology, which has been identified as a key mediator between maternal

Table 1
Domains of Parental Employment

Mothers' Job Characteristics

	NLSY	PSID	CPS
employment in years prior to birth of child	x	x	
employment hours during pregnancy	x	x	
length of maternity/parenting leave	x	x	
age of child when mother began (resumed) employment	x	x	x**
hours of work each quarter of first year	x	x	x**
hours of work each year of child's life	x	x	x**
summer hours (if different than rest of year)	x		x**
number of job changes each year	x	x	x**
weeks of unemployment (looking for work)	x	x	x
mother's satisfaction with schedule (subjective measure)			
salary - hourly and yearly	x	x	x
proportion of family income contributed			
record of fringe benefits received	x	x*	
paid vacation, health and dental insurance, maternity leave, flexible schedule	x	x*	
social security coverage on job			
whether employee experienced downsizing		x	
occupational complexity of job			
3 digit occupational code	x	x	x
job satisfaction	x		
peer relations	x*		
income	x*		
physical safety and cleanliness	x*		
preference for employment	x***		

x available every year of data collection
x* only available in one or occasional years
x** only data on individuals within the household. Data set is a household survey and individuals outside of the household can not be traced.
x*** for those who were currently unemployed only.
x**** will be available in 1994 for youth in their teens

Table 1 - continued
Fathers' Job Characteristics

	NLSY	PSID	CPS
father's hours of work each year of child's life	x**	x	x**
summer hours (if different)			x**
number of job changes each year			x**
weeks of unemployment (looking for work)	x**	x	x**
		x	
father's satisfaction with schedule (subjective measure)			
		x	
salary - hourly and yearly	x**	x	x**
	x*		
record of fringe benefits received		x*	
paid vacation, health and dental		x*	
insurance, maternity leave, flexible schedule			
social security coverage on job			
does father pay child support			
whether employee experienced downsizing		x	
occupational complexity of job			
3 digit occupational code	x**	x*	x*
job satisfaction			
peer relations			
income			
physical safety and cleanliness			

x available every year of data collection
x* only available in one or occasional years
x** only data on individuals within the household. Data set is a household survey and individuals outside of the household can not be traced.
x*** for those who were currently unemployed only.
x**** will be available in 1994 for youth in their teens

Table 2
Family and Child Care environment

Mother	NLSY	PSID	CPS
HOME scale	x		
Amount of time spent with child on typical weekday between 7 a.m. and 9 p.m.			
time spent with child on typical weekend day			
time spent with child during summer			
time spent on leisure time			
time spent with spouse			
time spent in housework per day		x	
strain/gains of work to parenting			
strain/gains of work to marriage			
satisfaction with parenting			
sex role attitudes	x*		
number of children in household	x	x	x
Father			
Amount of time spent with child on typical weekday between 7 a.m. and 9 p.m.			
time spent with child on typical weekend day			
time spent with child during summer			
time spent on leisure time			
time spent with spouse			
time spent in housework per day		x	
If non-custodial parent, number of hours spent with child during typical week			
strain/gain of work to parenting			
strain/gain of work to marriage			
satisfaction with parenting			
sex role attitudes			
number of children in the household		x	x

x available every year of data collection
x* only available in one or occasional years
x** only data on individuals within the household. Data set is a household survey and individuals outside of the household can not be traced.
x*** for those who were currently unemployed only.
x**** will be available in 1994 for youth in their teens

Table 2 continued
Family and Child care environment

	NLSY	PSID	CPS
longitudinal history of child care arrangements	x		
type of care - center, family day care, relative at home, relative at other's home	x		
type of after school care	x		
ratio of adult to child	x*		
caregiver's training	x		
caregiver's educational background			
number of changes over year	x		
number of child care arrangements in a week	x		

Table 3

Child Developmental Outcome Measures

	NLSY	PSID	CPS
Cognitive development	x		
Grade failure	x		
Educational grade achievement	x		
Socio-emotional development	x		
Behavior problems	x		
Attitude towards work	x****		
High School drop out	x****	x	x
Teenage Birth		x	x

Table 4

Parental Resources

Mother	NLSY	PSID	CPS
age	x	x	x
age at first birth	x	x	
marital status	x	x	x
verbal ability	x		
self esteem	x*		
depression	x*		
other adults in household	x	x	
social support from spouse			
social support from other family			
social support from friends, neighbors			
Father			
age		x	x**
age at first birth			
marital status			
living with child		x	x
verbal ability		x*	
educational attainment	x**	x	x**
depression			
social support from spouse			
social support from friends			
paying child support to other children - amount	x**		x

x available every year of data collection
x* only available in one or occasional years
x** only data on individuals within the household. Data set is a household survey and individuals outside of the household can not be traced.
x*** for those who were currently unemployed only.
x**** will be available in 1994 for youth in their teens

CONCLUSION

Decisions about data collection on parental employment needs to be considered within the context of several significant demographic shifts which may affect the well-being of children. (i) Labor force participation of women with children has increased dramatically --

employment and child development, is only measured in the NLSY and even here this is measured only at one point in time.

CONCLUSION

Decisions about data collection on parental employment needs to be considered within the context of several significant demographic shifts which may affect the well-being of children. (i) Labor force participation of women with children has increased dramatically -- more women are entering and remaining in the labor force than ever before. The largest recent rise in labor force participation is among women with children under one year-old. These trends are necessitating changed family relationships. For married women, a two-earner family is now the norm, rather than the so-called "traditional" family of a male breadwinner and an at- home mother. Among the growing numbers of single-parent female-headed households who are also employed, there is the stress of managing employment schedules and child care arrangements without the support of a spouse. (ii) For women, a job, rather than marriage, is the institution which promises economic security via the provision of fringe benefits that include health insurance, pensions and social security coverage. (iii) Inequities between professional workers and less-skilled workers are increasing. While workers with higher schooling levels and more experience have been able to keep up with inflation, the real earnings of younger and less educated workers have fallen sharply. (iv) Poverty among children is increasing when parents are out of the labor force and dependent on public transfers, primarily AFDC. (v) Finally, poor women with young children, who are out of the labor force and receiving AFDC, are now expected to obtain employment. Women with children three years-old and older are now considered eligible for employment and must participate in training and employment programs.

A serious gap in the research on parental employment and child well-being are studies that focus on the experience of low-income and minority families who are employed. The majority of research on maternal employment has focused on homogenous samples - primarily white, two-parent, middle-class families. The work experience of a parent in a low-paying unskilled job with little opportunities for self-direction or promotion are obviously very different than that of a parent working in a professional or managerial job. Very little is known about the working conditions of low-income workers and the effects of this employment situation on their children. In 1992, women earned only 71% of the wages earned by men. Women of color experience the most severe pay inequities with black women earning 64 cents, Hispanic women 55 cents and white women 70 cents for each dollar earned by a white man. Men of color also experience significant wage discrimination. A large part of the wage gap is due to the fact that women and people of color tend to work in technical/sales or service occupations where wages are low (National Committee on Pay Equity, 1994). Yet how these particular working conditions impact on parenting capacities has not been adequately examined. The link between early employment and developmental outcomes for the children of poor, single, and minority mothers remains largely unexplored. It is reasonable to assume that these mothers experience excessive demands in the family role, especially the role of single mother, in addition to coping with the negative effects of financial strain and unstable employment opportunities, all of which might have negative consequences for their children, both contemporaneously and over time (Dodge, Pettit, Bates, 1994; Downey & Coyne, 1990; Huston, McLoyd & Garcia-Coll, 1994; McLoyd, 1990, 1994; Jackson, 1992, 1993, 1994; Leadbeater & Linares, 1992).

As welfare reform policies increasingly include mandated maternal employment, research is needed on the impact of low-income working women's employment conditions on the women themselves and on their children. Research is needed which investigates the effects on child well-being if a family moves from a "below poverty level" to a "near poverty level" of family income. Brooks-Gunn and Smith (1994) found that the cognitive abilities of children in families who left AFDC but remained poor were lower than those children whose families remained on public assistance.

In summary, our review of the needs for monitoring parental employment has suggested that there are three categories of employment that can be monitored now from existing data. We recommend that each of these areas be looked at separately for two-parent families, mother-only families and father-only families. We also suggest that analyses be conducted looking at these families by socio-economic groupings using cross-tab analyses by ethnic/racial groups and by the mother's educational achievement. Our review suggests that the top priority in gathering social indicator data on parental employment it to begin immediately with data from Current Population Surveys with new analyses done with breakdowns by age of child. The second and third levels described in our paper (family environment and child outcomes) are also important and available in SIPP, NLSY or PSID. However, with time and budgetary constraints we suggest that the first level be the highest priority of indicators monitoring in this area.

References

Ainsworth, M. (1964). Patterns of attachment behavior shown by the infant in interaction with his mother. *Merrill-Palmer Quarterly of Behavior and Development, 10*, 51-58.

Baker, P., & Mott, F. (1989). *NLSY handbook 1989: A guide and resource document for the National Longitudinal Survey of Youth 1986 child data*. Columbus, OH: Center for Human Resource Research, The Ohio State University.

Bane, M.J., & Ellwood, D. (1983). *Slipping into and out of poverty: The dynamics of spells*. NBER working paper #1199. Cambridge, MA: National Bureau of Economic Research.

Barling, J & Fullagar, C., & Marchl-Dingle, J. (1986). Employment commitment as a moderator of the maternal employment status/child behavior relationship. *Journal of Occupational Behavior, 9*, 113-122.

Barnett, R.C. & Marshall, N.C. (1992). Worker and mother roles, spillover effects, and psychological distress. *Women and Health, 18*(2), 9-14.

Baruch, G.K., & Barnett, R.C. (1987). Role quality and psychological well-being. In F. Crosby (Ed.), *Spouse, parent, worker* (pp. 63-84). New Haven, CT: Yale University Press.

Baydar, N., & Brooks-Gunn, J. (1991). Effects of maternal employment and child-care arrangements in infancy on preschoolers' cognitive and behavioral outcomes: Evidence from the children of the NLSY. *Developmental Psychology, 27*, 918-931.

Belsky, J. (1988) The "effects" of infant day care reconsidered. *Early Childhood Research Quarterly, 3*, 227-235.

Belsky, J., & Eggebeen, D. (1991). Early and extensive maternal employment/child care and 4-6 year olds' socioemotional development: Children of the National Longitudinal Study of Youth. *Journal of Marriage and the Family, 53*, 1083-1099.

Belsky, J., & Steinberg, L. (1978). The effects of day care: A critical review. *Child Development, 49*, 706-717.

Bergmann, B. (1986). *The economic emergence of women*. New York: Basic Books.

Blau, F., & Grossberg, A. (1990). *Maternal labor supply and children's cognitive development*. Unpublished Working Paper #3536. National Bureau of Economic Research.

Bornstein, M. (1994, in press). *Handbook of parenting*. Hillsdale, NJ: Erlbaum.

Bowlby, J. (1951). *Maternal care and mental health*. Geneva: World Health Organization.

Bowlby, J. (1969). *Attachment and loss: Vol. 1*. New York: Basic Books.

Bradley, R., Caldwell, B., Rock. S., Hamrick, H. & Harris, P. (1988). Home observation for measurement of the environment: Development of a home inventory for use with families having children 6 to 10 years old. *Contemporary Educational Psychology, 13*, 58-71.

Bradley, R. & Tedesco, L. (1982). Environmental correlates of mental retardation. In J. Lachenmeyer & M. Gibbs (Ed.), The psychology of the abnormal child (pp. 155-188). New York: Gardner Press.

Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist, 34*, 844-850.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology, 22*, 723-742.

Bronfenbrenner, U., & Crouter, A. (1982). Work and family through time and space. In S.B. Kamerman & C.D. Hayes (Eds.), *Families that work: Children in a changing world* (pp. 39- 83). Washington, DC: National Academy Press.

Brooks-Gunn, J. (1994, in press). Growing up poor: Context, risk and continuity in the Bronfenbrenner tradition. In P. Moen, G.H. Elder, & K. Lusher, (Eds.), *Linking lives and contexts: Perspective on the ecology of human development.*

Brooks-Gunn, J., Brown, B., Duncan, G. D., Moore, K.A. (1994). *Child Development in the Context of Family and Community Resources: An Agenda for National Data Collections.* (Paper for NAS Conference on "Integrating Federal Statistics on Children").

Brooks-Gunn, J., Klebanov, P., Liaw, F., & Spiker, D. (1993). Enhancing the development of low birthweight, premature infants: Changes in cognition and behavior over the first three years. *Child Development, 64,* 736-753.

Brooks-Gunn, J. & Smith, J. (1994, manuscript in preparation). Transitions in and out of welfare: Effects on children.

Bumpass, L.L. & Sweet, J.A. (1980). *Patterns of employment before and after childbirth.* Series 23 (4). Washington DC: US Department of Health, Education and Welfare.

Chase-Lansdale, P. L., Mott, F. L., Brooks-Gunn, J., & Phillips, D. (1991). Children of the NLSY: A unique research opportunity. *Developmental Psychology, 27*(6), 918-931.

Chase-Lansdale, P.L. & Owen, M.T. (1987). Maternal employment in a family context: Effects on infant-mother and infant-father attachments. *Child Development, 58,* 1505-1512.

Chodorow, N. 91978). *The reproduction of mothering: Psychoanalysis and the sociology of gender.* Berkeley, CA: University of California Press.

Clarke-Stewart, K.A. (1988). The 'effects' of infant day care reconsidered. *Early Childhood Research Quarterly, 3*, 293- 318.

Clarke-Stewart, K.A. (1989). Day care: Maligned or malignant. *American Psychologist, 44*, 266-273.

Cleary, P.D., & Mechanic, D. (1983). Sex differences in psychological distress among married people. *Journal of Health And Social Behavior, 24*, 111-121.

Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*, S95-S120.

Coleman, J. S. (1991). *Policy perspectives: Parental involvement in education.* Washington, DC: Office of Educational Research and Improvement: U.S. Department of Education.

Conger, R., Conger, K., Elder, G., Lorenz, F., Simons, R., & Whitebeck, L. (1992). A family process model of economic hardship and adjustment of early adolescent boys.

Cott, N. (1977). *The bonds of motherhood.* New Haven: Yale University Press.

Crouter, A.C., & McHale, S.M. (1993). Temporal rhythms in family life: Seasonal variation in relation between parental work and family processes. *Developmental Psychology, 29*(2), 198-205.

Danziger, S., & Gottschalk, P. (1990). *How have families with children been faring?* IRP Discussion Paper #801-86. Madison: University of Wisconsin.

Darling-Fisher, C.S. & Tiedje, L.B. (1990). The impact of maternal employment characteristics on father's participation in child care. *Family Relations, 39*(1), 20-26.

DeMause, L. (1974). *The history of childhood.* New York: Psychohistory Press.

Desai, S., Chase-Lansdale, L., & Michael, R., (1989). Mother or market?: Effects of maternal employment on cognitive development of four year old children. *Demography, 26*, 545-561.

Dinnerstein, D. (1976). *The mermaid and the minotaur: Sexual arrangements and human malaise.* New York: Harper and Row.

Dodge, K.A., Pettit, G.S., & Bates, J.E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development* [Special issue: Children and Poverty], *65*(2), 649-665.

Downey, G., & Coyne, J. (1990). Children of depressed parents: An integrative review. *Psychological Bulletin, 108*, 50-76.

Drucker, P. (1976).*The unseen revolution: How pension fund socialism came to America.* New York: Harper & Row.

Dunn, L. M. & Dunn, L. M. (1981). *Peabody picture vocabulary test-revised.* American Guidance Service.

Eggebeen, D. & Lichter, D. (1991). Race, family structure, and changing poverty among American children. *American Sociological Review, 56*, 801-817.

Ehrenreich, B. & English, D. (1978). *For her own good: 150 years of expert advise to women.* Garden City, NY: Doubleday.

Elder, G., Liker, J., & Cross, C. (1984). Parent-child behavior in the Great Depression: Life course and intergenerational influences. In P. Baltes & O. Brim (Eds.), Life span development and behavior (Vol. 6, pp. 109-158).

Elder, G., Nguyen, T., & Caspi, A. (1985). Linking family hardship to children's lives. Child Development, 56, 361-375.

Erickson, M., Sroufe, A.,& Egeland, B. (1985). The relationship between quality of attachment and behavior problems in preschool children in a high-risk sample. In I. Bretherton & E. Waters (Eds.), *Growing points in attachment theory and research. Monographs of the Society for Research in Child Development, 50*(1-2, Serial No. 209), 147-166.

Farel, A. (1980). Effects of preferred maternal roles, maternal employment and socio-demographic status on school adjustment and competence. *Child Development, 50*, 1179-1186.

Feree, M.M. (1990). Beyond separate spheres: Feminism and family research. *Journal of Marriage and the Family, 52*, 866-884.

Forehand, R., McCombs, A., & Brody, G. (1987). The relationship between parental depressive mood states and child functioning. *Behavior Research and Therapy*, *9*, 1-20.

Fraiberg, S. 91977). *Every child's birth right: In defense of mothering*. NY: Basic Books.

Galinsky, E. Howes, C., Kontos, S. & Shinn, M. (1994). The study of children in family child care and relative care. Families and Work Institute: New York.

Garmezy, N. & Rutter, M. (Eds.) (1983). *Stress, coping, and development in children*. New York: McGraw-Hill.

Gilbert, L. (1985). Men in dual-career families: Current realities and future prospects. Hillsdale, NJ: Erlbaum.

Glendon, M.A. (1981). *The new family and new property*. Seattle, W: Butterworths.

Goode, W.J. (1960). A theory of role strain. *American Sociological Review*, *25*, 483-496.

Gottfried, A.E. & Gottfried, A.W. (Eds.). (1988). *Maternal employment and children's development: Longitudinal research*. New York: Plenum Press.

Gottfried, A.E. & Gottfried, A.W. (1994). Role of maternal and dual-earner employment status in children's development. In A. E. Gottfried & A.W. Gottfried (Eds.), *Redefining Families: Implications for children's development*. Plenum Press: New York.

Gove, W.R., & Peterson, C. (1980). Update of the literature on personal and marital adjustment: The effect of children and the employment of wives. *Marriage and Family Review*, *3*, 63-69.

Gove, W., & Zeiss, C. (1987). Multiple roles and happiness. In F. Crosby (Ed.), *Spouse, parent, worker*. New Haven: Yale University Press.

Guidubaldi, J., & Nastasi, B.K., (1987). *Home environment factors as predictors of child adjustment in mother-employed households: Results from a nationwide study*. Paper presented at the biennial meeting of the Society of Research in Child Development, Baltimore, MD.

Harris, B. (1979). Careers, conflict and children: The legacy of the cult of domesticity. In A. Roland & B. Harris (Eds.), *Career and motherhood: Struggles for a new identity* (pp. 55-86). New York: Human Sciences Press.

Haskins, R. (1985). Public school aggression among children with varying day-care experience. *Child Development*, *56*, 689-703.

Haveman, R. & Wolfe, B. (1991). *On the determinants of children's well being and economic success: A policy perspective*. Paper prepared for APPAM meetings, October.

Haveman, R. & Wolfe, B. (1994). *Do investments in children matter?: A review of methods and findings*. Mimeo: University of Wisconsin.

Hochschild, A. (1989). *The second shift*. New York: Viking.

Hock, E. (1980). Working and nonworking mothers and their infants: A comparative study of maternal care-giving characteristics and infant social behavior. *Merrill Palmer Quarterly*, *26*, 79-101.

Hofferth, S.L., Brayfield, A., Deich, S., & Holcomb, P. (1991). *National Child Care Survey*, 1991. Urban Institute Report 91-5. Washington, DC: Urban Institute Press.

Hoffman, L.W. (1974). Fear of success in males and females: 1965 and 1971. *Journal of Consulting and Clinical Psychology, 42*, 353-358.

Hoffman, L.W. (1979). Maternal employment: 1979. *American Psychologist, 34*, 859-865.

Hoffman, L.W. (1989). Effects of maternal employment in the two-parent family. *American Psychologist, 44*, 283-292.

Huston, A.C., McLoyd, V.C., & Garcia-Coll, C. (1994). Children and poverty: Issues in contemporary research. *Child Development, 65*, 275-282.

Jackson, A.P. (1992). Well-being among single, black employed mothers. *Social Service Review, 66*, 399-409.

Jackson, A.P. (1993). Black single working mothers in poverty: Preferences for employment, well-being, and perceptions of preschool-age children. *Social Work, 38*, 26-34.

Jackson, A.P. (1994). The effects of role strain on single, working, black mothers' perceptions of their young children. *Social Work Research, 18*, 36-40.

Jahoda, M. (1982). *Employment and unemployment: A social-psychological analysis*. Cambridge: Cambridge University Press.

Jouriles, E., Murphy & O'Leary. (1980). Effects of maternal mood on mother-son interaction patterns. *Journal of Abnormal Child Psychology, 17*, 513-526.

Juster, F.J. & Stafford, F. (Eds.). (1985). *Time, goods and well-being*. Ann Arbor, MI: University of Michigan.

Kamerman, S.B. (1980). *Parenting in an unresponsive society: Managing work and family life*. New York: Free Press.

Kamerman, S.B. (1988). Maternity and parenting benefits: An international review. In E. Zigler and O. Frank (Eds.), *The parental leave crisis: Toward a national policy*. New Haven: Yale University Press.

Kamerman, S.B., & Hayes, C.D. (Eds.) (1982). *Families that work: Children in a changing world*. Washington, DC: National Academy Press.

Kamerman, S.B., & Kahn, A. (1981) *Child care, family benefits and working parents*. New York: Columbia University Press.

Kamerman, S.B., & Kahn, A. (1988) *Mothers alone: Strategies for a time of change*. Dover, MA: Auburn.

Kamerman, S.B. & Kahn, A. (1991). *Child care, parental leave, and the under 3's*. New York: Auburn House.

Kessler, R., & McRae, J. (1982). The effects of wives' employment on the mental health of married men and women. *American Sociological Review, 47*, 216-227.

Kessler-Harris, A. (1982). *Out to work: A History of wage-earning women in the United States*. Oxford: Oxford University.

Kisker, E., Hofferth, S., Phillips, D. & Faruhar, E. (1991). *A profile of child care settings: Early education and care in 1990*. Washington, DC: U.S. Government Printing Office.

Kohn, M. (1969). *Class and conformity: A study in values*. Chicago: University of Chicago Press.

Kohn, M., & Schooler, C. (1973). Occupational experience and psychological functioning: An assessment of reciprocal effects. *American Sociological Review, 38*, 97-118.

Kohn, M., & Schooler, C. (1982) Job conditions and personality: A longitudinal assessment of their reciprocal effects. *American Journal of Sociology, 87*, 1257-1286.

Krein, S. & Beller, A. (1988). Educational attainment of children from single-parent families: Differences between exposure, gender and race. Demography, 25, 221-234.

Leadbeater, B.J., & Linares, O. (1992). Depressive symptoms in Black and Puerto Rican adolescent mothers in the first 3 years post-partum. *Development and Psychopathology, 4*, 451-468.

Leibowitz, A. (1974). Education and home production. *American Economic Review, 64*, 243-250.

Lerner, J., & Galambos, N. (1985). Maternal role satisfaction, mother-child interaction, and child temperament: A process model. *Developmental Psychology, 21*, 1157-1164.

Lewin, T. (1994, October 15). Working women say bias persists. *The New York Times*, pp. B1.

Lopata, H., Miller, C., & Barnewolt, D. (1984). *City women: Work, jobs, occupations careers*. New York: Praeger.

Maccoby, E. E., & Martin, J. A. (1983). Socialization in the context of the family: Parent-child interaction. In P. H. Mussen & E. M. Hetherington (Eds.), *Handbook of child psychology: Socialization, personality, and social development* (pp. 1-102). New York: John Wiley & Sons.

Manke, B., Seery, B. L., & McHale, S.M. (1994). The three corners of domestic labor: Mothers', fathers', and children's weekday and weekend housework. *Journal of Marriage and the Family, 56*(3), 657-668.

Marks, S.R. (1977). Multiple roles and role strain: Some notes on human energy, time, and commitment. *American Sociological Review, 42*, 921-936.

Marshall, N. & Barnett, R. (1991). Race, class and multiple role strains and gains among women employed in the service sector. *Women and Health, 17*, 1991.

McHale, S., Bartko, W., Crouter, A., & Perry-Jenkins, M. (1990). Children's housework and psychosocial functioning: The mediating effects of parents' sex role behaviors and attitudes. *Child Development, 61*, 1413-1426.

McHale, S.M. & Crouter, A.C. (1992). You can't always get what you want: Incongruence between sex-role attitudes and family work roles and its implications for marriage. *Journal of Marriage and the Family, 54*, 537-547.

McLanahan, S. (1985). Family structure and the reproduction of poverty. *American Journal Of Sociology, 90*, 873-901.

McLanahan, S. & Sandefur, G. (1994).

McLoyd, V.C. (1990). The impact of economic hardship on black families and children: Psychological distress, parenting, and socioemotional development. *Child Development, 61*, 311-346.

Menaghan, E., & Parcel, T. (1991). Determining children's home environments: The impact of maternal characteristics and current occupational and family conditions. *Journal of Marriage and the Family, 53*, 417-431.

Miller, J., Schooler, C., Kohn, M., & Miller, K. (1979). Women and work: The psychological effects of occupational conditions. *American Journal of Sociology, 85*, 66-94.

Moen, P., & Dempster-McClain, E. (1987). Employed parents: Role strain, work time and preferences for less. *Journal of Marriage and the Family, 49*, 579-590.

Moore, K., & Hofferth, S. (1979). Effects of women's employment on marriage. *Marriage and Family Review, 2*, 27-36.

Morgan, G. (1987). *The national state of child care regulation*. Watertown, MA: Work/Family Directions.

Mortimer, J.T., & Sorenson, G. (1984). Men, women, work, and family. In K.M. Borman, D. Quarm, & S. Gideonese (Eds.), *Women in the workplace* (pp. 139-167). Norwood, NJ: Ablex.

National Committee on Pay Equity. (1994). *Face the facts*. Mimeograph, Washington DC.

Oakley, A. (1974) *Women's work: A history of the housewife*. New York: Pantheon Books.

Oppenheimer, V. (1979) *The female labor force in the United States: demographic and economic factors governing its growth and changing composition*. Westport CT: Greenwood Press.

Parcel, T. (1989). Comparable worth, labor markets and earnings. In R. Michael, H. Hartmann & B. O'Farrell (Eds.), *Pay equity: Empirical inquiries*, Washington, DC: National Academy Press.

Parcel, T. & Menaghan, E. (1994). Early parental work, family social capital, and early childhood outcomes. *American Journal of Sociology, 99*, 972-1009.

Parke, R.D. (1982). The father-infant relationship: A family perspective. In P. Berman (Ed.), *Women: A developmental perspective*. (pp. 349-364).

Parry, G. (1986). Paid employment, life events, social support, and mental health in working-class mothers. *Journal. of Health and Social Behavior, 27*, 193-208.

Piotrkowski, C., & Katz, M. (1982). Indirect socialization of children: The effects of mothers' jobs on academic behaviors. *Child Development, 53*, 1520-1529.

Pleck, J. (1984). The work-family role system. In P. Voydanoff (Ed.), *Work and family*. Palo Alto, CA: Mayfield.

Pleck, J.H. (1985). *Working wives/working husbands*. Beverly Hills, CA: Sage.

Pleck, J.H.(1986). Employment and fatherhood: Issues and innovative policies. In M. Lamb (Ed.), *The father's role*. New York: Wiley.

Pleck, J.H., Staines, G.L., & Lang, L. (1980). Conflicts between work and family life. *Monthly Labor Review, 103*, 29-31.

Presser, H. (1989) Can we make time for children? The economy, work schedules, and child care. *Demography, 26*, 524-541

Radloff, L. (1977). Sex differences in depression: The effects of occupational and marital status. *Sex Roles, 1*, 249-265.

Rallings, E.M., & Nye, F.I. (1979). Wife-mother employment, family and society. In W.R. Burr (Eds.), *Contemporary theories about the family: Vol. 1.* New York: Free Press.

Rauh, V. (1994, October). Employment as social capital for pregnant women. Paper presented at the 122nd Annual Meeting of American Public Health Association; Public Health and Diversity: Opportunity for Equity, Washington, D.C.

Reich, C. (1964). The new property. *Yale Law Journal (April)*.

Repetti, R. (1987). Linkages between work and family roles. In S. Oskamp (Ed.), *Family processes and problems: Social psychological aspects* (pp. 98-127). Newbury Park, NJ: Sage.

Ross, C., Miroskwy, J., & Huber, J. (1983). Dividing work, sharing work and in-between: Marriage patterns and depression. *American Sociological Review, 48*, 809-823.

Scarr, S., Phillips, D., & McCartney, K. (1990). Facts, fantasies and the future of child care in the United States. *Psychological Science, 1*, 26-35.

Schwartz, J.C., Strickland, R.G., & Krolick, G. (1974). Infant day care: Behavioral effects at preschool age. *Developmental Psychology, 54*, 502-506.

Schwartz, P. (1983). Length of day-care attendance and attachment behavior in eighteen-month-old infants. *Child Development, 54*, 1073-1078.

Sieber, S.D. (1974). Toward a theory of role accumulation. *American Sociological Review, 39*, 567-578.

Silverstein, L. (1991). Transforming the debate about child care and maternal employment. *American Psychologist, 46*: 1025-1032.

Smith, J. (1994). *Maternal employment and the young child.* Unpublished doctoral dissertation. Columbia University, New York.

Solberg, E. & Laughlin, T. (1994, in press). The gender pay gap, fringe benefits and occupational crowding. *International and Labor Relations Review.*

Spitze, G. (1988).Women's employment and family relations: A review. *Journal of Marriage and the Family, 50*, 595-618.

Staines, G. (1980). Spillover versus compensation: A review of the literature on the relationship between work and nonwork. *Human Relations, 33*, 111-129.

Timmer, S., Eccles, J., & O'Brien, K. (1985). How children use time. In F.J. Juster & F. Stafford (Eds.), *Time, goods and well-being.* Ann Arbor, MI: University of Michigan.

U.S. Bureau of Labor Statistics. (1990, March). *Monthly labor review.* Washington, DC: Government Printing Office.

U.S. Department of Census. (1990). *Current population reports.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Census. (1991). *Current population reports*. Washington, DC: U.S. Government Printing Office.

Vaughn, B., Deane, K. & Waters, E. (1985). the impact of out-of-home care on child-mother attachment quality: Another look at some enduring questions. *Monographs of the Society for Research in Child Development, 50*, (1-2, Serial NO. 209), 110-135.

Vaughn, B., Gove, F., & Egeland, B. (1980). The relationship between out-of-home care and the quality of infant-mother attachment in an economically disadvantaged population. *Child Development*. *51*, 1203-1214.

Wachs, T.D. & Gruen, G. (1982). Early experience and human development. New York: Plenum.

Walker, L., & Best, M. (1991). Well-being of mothers with infant children: A preliminary comparison of employed women and homemakers. *Women and Health*, *17*, 71-87.

Warr, P., & Parry, G. (1982). Paid employment and women's psychological well-being. *Psychological Bulletin*, *91*, 498-516.

Woods, M. (1972). The unsupervised child of the working mother. *Developmental Psychology*, *6*, 14-25.